REPORT RESUMES

ED 003 826 24
A PROJECT TO STUDY THE NATURE OF EFFECTIVE PHYSICS TEACHING.
BY- SNIDER, RAY H.
CORNELL UNIV., ITHACA, N.Y.
REPORT NUMBER CRP-S-280 PUB DATE 65
EDRS PRICE MF-$0.54 HC-$13.64 341P.

DESCRIPTORS- PHYSICS, PHYSICS CURRICULUM, *PHYSICS INSTRUCTION, HIGH SCHOOLS, VERBAL COMMUNICATION, *STUDENT TEACHER RELATIONSHIP, *CLASSROOM RESEARCH OBSERVATION, *EFFECTIVE TEACHING, *ACADEMIC ACHIEVEMENT, ITHACA, NEW YORK, FLANDERS METHOD

THE FLANDERS METHOD OF INTERACTION ANALYSIS WAS EMPLOYED IN AN ANALYTICAL STUDY OF HIGH SCHOOL PHYSICS TEACHING. THE STUDY OBJECTIVES WERE (1) TO DESCRIBE THE CHARACTERISTICS AND PATTERNS OF TEACHER-STUDENT VERBAL INTERACTION IN HIGH SCHOOL CLASSROOMS, (2) TO DETERMINE RELATIONSHIPS BETWEEN TEACHER EFFECTIVENESS, AS MEASURED BY SELECTED ASPECTS OF STUDENT ACHIEVEMENT, AND FLEXIBILITY OF TEACHER INFLUENCE, AND (3) TO PROVIDE A STORE OF USEFUL DATA ON THE DISCOURSE OF HIGH SCHOOL PHYSICS CLASSES WHICH WILL FACILITATE FURTHER ANALYSES. CLASSES WERE OBSERVED FOR 4 CONSECUTIVE DAYS IN THE FALL AND 3 CONSECUTIVE DAYS IN THE SPRING (1964-65). DATA GATHERED DURING THESE PERIODS WERE ANALYZED USING A CONTROL DATA 1604 COMPUTER. IT WAS CONCLUDED THAT (1) THE FLANDERS METHOD OF INTERACTION ANALYSIS WAS A FEASIBLE METHOD FOR ACQUIRING USEFUL DATA CONCERNING WHAT TAKES PLACE IN THE HIGH SCHOOL PHYSICS CLASSROOM, (2) THERE IS LITTLE OR NO TEACHING TO CULTIVATE STUDENT INQUIRY IN THE HIGH SCHOOL PHYSICS LABORATORY, AND (3) STUDENT GAINS WERE FOUND IN THIS STUDY TO BE LIMITED TO SUBJECT MATTER FACTS AND CONCEPTS. SINCE UNDERSTANDINGS OF THE NATURE OF SCIENCE (PROCESS, METHODS, AIMS) ARE CONSIDERED IMPORTANT, THE LACK OF GAIN ON THIS TEST RAISES SERIOUS QUESTIONS ABOUT THE CURRICULUM AND PHYSICS TEACHING. (LP)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A PROJECT TO STUDY THE NATURE OF EFFECTIVE PHYSICS TEACHING

Cooperative Research Project No. S-280

Ray Merrill Snider

Cornell University
Ithaca, New York

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

This research project was reported in a dissertation submitted to Cornell University and entitled: "A PROJECT TO STUDY THE NATURE OF PHYSICS TEACHING USING THE FLANDERS METHOD OF INTERACTION ANALYSIS."

ED 003826

PRECEDING PAGE BLANK-NOT FILMED

BIOGRAPHICAL SKETCH

The author was born in New York City on June 16, 1930, son of Forrest Ray Snider and Mabel V. Snider Shenot. He attended Amityville (New York) High School and graduated from Milford (New York) Central High School in June 1948.

He served four years in the United States Air Force, the latter three of which were spent in Germany as a radio operator. He was honorably discharged in May 1952 with the rank of Staff-Sergeant.

He obtained the Associate in Arts Degree in Mathematics from Orange County Community College, Middletown, New York in 1952. In 1954 he earned a Bachelor of Science in Mathematics at State University of New York at Albany; the Master of Science in Natural Science was received in 1960 from Rensselaer Polytechnic Institute, Troy, New York. In addition, the author studied and taught at Oklahoma State University, Stillwater during 1961-1962.

He taught secondary chemistry, physics, and mathematics at F. D. Roosevelt High School, Hyde Park (New York) Central School District from 1956 through 1961. During these latter three years, he chaired the Department of Science.

He was admitted to Cornell University in June 1961 and completed requirements for the Ph.D. in November 1965, with a major in Science Education and minors in Physics and Mathematics.

He has held membership in the honorary societies Kappa Mu Epsilon (mathematics), Sigma Pi Sigma (physics), and Phi Delta Kappa (education). He is also a member of Phi Kappa Phi. In addition, he holds membership in the New York State Science Teachers Association, the National Science Teachers Association, the National Association for Research in Science Teaching, and the American Association of Physics Teachers.

He married Frances J. Triolo in 1953; they have two children, Catherine Ann Snider, born May 5, 1963 and Matthew Ray Snider, born December 17, 1964.

ACKNOWLEDGEMENTS

The research reported herein would not have been possible without the support and assistance of many individuals and organizations. It was carried out with funds contributed by Cornell University and through a research grant provided by the Small Contract Program, Basic Research Branch, U. S. Office of Education (Project S-280). The author gratefully acknowledges Dr. Frederick H. Stutz, Dean of the School of Education; Dr. W. Keith Kennedy, Director of Research; and Dr. Joseph F. Metz, Jr., Assistant Director of Research, all of Cornell University; as well as Dr. Howard H. Cummings, Research Coordinator of the U. S. Office of Education. In addition, the author is particularly appreciative of the teachers, students, and school administrators who participated in the study (anonymity precludes their specification). Their fine spirit of willingness and cooperation was outstanding throughout.

The author is especially indebted to his Graduate Committee for their guidance and support throughout his years of graduate study at Cornell University. An expression of deep gratitude is extended to his Committee Chairman and Project Sponsor, Dr. Philip G. Johnson, Professor of Science Education and Chairman of the Division of Science Education, for a memorable association that will be of lasting value. Special thanks are due Committee members

Dr. Kenneth I. Greisen, Professor of Physics; Dr. Alex Rosenberg, Professor of Mathematics; and Dr. Matthew H. Bruce, Jr., Assistant Professor of Science Education. The author is particularly grateful for the numerous helpful suggestions and criticisms provided by Dr. Bruce as the research progressed.

It is a pleasure to acknowledge Dr. Helen L. Wardeberg, Dr. Mauritz Johnson, Jr., Dr. Marvin D. Glock, and Dr. Jason Millman, all of the School of Education, Cornell University, for their reviews of the original proposal. He also thanks Dr. Ned A. Flanders, School of Education, The University of Michigan, Ann Arbor, for valuable communications.

The author has enjoyed his associations and stimulating discussions with his fellow graduate students, especially Dr. Jal S. Parakh, Mr. Lawrence E. Wightman, and Mr. Charles C. Matthews. He is grateful, too, for the excellent assistance of Mrs. Rita Collier who typed the thesis and carried out numerous other tasks.

Finally, the author feels incapable of adequately expressing his deep appreciation of his wife, Frances J. Snider, who so lovingly provided her support and encouragement.

TABLE OF CONTENTS

Page

BIOGRAPHICAL SKETCH ii

ACKNOWLEDGEMENTS iv

LIST OF TABLES x

LIST OF FIGURES xiv

CHAPTER ONE: THE RESEARCH PROBLEM 1

- Introduction 1
- Scope of the Problem 1
- Purpose of the Study 4

CHAPTER TWO: REVIEW OF RELATED LITERATURE 6

- Introduction 6
- The Need for Systematic Observation 7
- Studies of Teacher Behavior Employing Systematic Observation 9
- Studies on Science Teaching 21
- The Work of *Flanders* 25
- Some Theoretical Constructs of Teaching 31
- Concluding Remarks 33

CHAPTER THREE: THE EXPERIMENTAL DESIGN 35

- Objectives 35
- Procedures 36
  - General Design 36
  - Population and Sample 40
  - Analyses 40
- Concluding Remarks 42

CHAPTER FOUR: THE SAMPLE OF PARTICIPATING TEACHERS AND SCHOOLS 43

- Introduction 43
- Sample Size 43
- The Sampling Technique Used 45
  - Reasons for Non-Random Sampling 45
  - The Population Defined 46
  - Selected Population and Sample Characteristics 47
  - School and Teacher Solicitation Procedure 51
- Comparing the Sample to the Population 52
- Concluding Remarks 55

Page

CHAPTER FIVE: SYSTEMATIC OBSERVATION IN THE HIGH SCHOOL PHYSICS CLASSROOM - COLLECTION OF DATA ... 58

- The Observation Instrument ... 58
  - Flanders' Categories ... 58
  - General Ground Rules Used ... 69
  - Special Ground Rules Used ... 70
- Observer Training Program ... 74
- Observation Techniques ... 75
  - Schedule of Classroom Visits ... 75
  - Observation Process in the Classroom ... 76
- The Effect of the Observer's Presence in the Classroom ... 79
- Major Activities in the Classroom ... 82
- Coding the Discourse ... 85
- Observer Reliability ... 87

CHAPTER SIX: TEACHER-STUDENT VERBAL INTERACTION IN THE HIGH SCHOOL PHYSICS CLASSROOM ... 95

- Introduction ... 95
- The Flanders Interaction Matrix ... 96
  - How the Matrix is Plotted ... 96
  - Areas of the Matrix ... 97
  - Selected I-D Percentages ... 103
  - Selected Matrix Columns, Cells, and Cell Groups ... 105
- The Wightman Program for Interaction Analysis ... 109
- Interaction Matrices of High School Physics Teaching ... 112
  - Time in Major Activities ... 112
  - The Grand Matrix of Regents Physics Teaching . 115
  - Grand Matrices of Major Activities ... 122
  - Further Comparisons of Interaction Matrices .. 130
- The Interaction Matrix as an Indicator of Teacher Style ... 138
  - Consistency of Matrices - A Rationale ... 138
  - The Darwin Chi-Square Tests ... 139
- Teacher Influence and Student Attitude ... 148
- Concluding Remarks ... 154

CHAPTER SEVEN: DETERMINATION OF TEACHER EFFECTIVENESS . 157

- Introduction ... 157
- Student Testing ... 158
  - Major Objectives Selected and Tests Used ... 158
  - Validity and Reliability of Tests ... 161
  - Procedures of Test Administrations ... 162
  - Reliability Determined From Test Results ... 164

Page

Student Achievement in Regents Physics 167
  Initial Ability and Previous Achievement 167
  Selection of Control Variables 174
  Regression Analyses 180
Teacher Effectiveness 190

CHAPTER EIGHT: TEACHER INFLUENCE AND EFFECTIVENESS 202

Introduction 202
Variation of Effectiveness Scores 202
The Mann-Whitney U Tests 209
  Indirectness of Teacher Influence and Effectiveness 212
  Flexibility of Teacher Influence and Effectiveness 218
  Selected Interaction Matrix Measures and Effectiveness 222
Concluding Remarks 231

CHAPTER NINE: IMPLICATIONS OF MAJOR FINDINGS 233

Implications Concerning the Usefulness of Interaction Analysis 233
Implications Concerning Directness of Teacher Influence 235
  The Normative Data 235
  The Laboratory Activity 240
Implications Concerning Teacher Effectiveness ... 242
Implications Concerning Further Research 245

CHAPTER TEN: SUMMARY 248

Objectives 248
Procedures 249
Major Findings 252
  Techniques 252
  Teacher Influence in the Classroom 252
  Teacher Influence in the Laboratory Activity . 254
  Teacher Influence and Effectiveness 254
Further Research 255

LITERATURE CITED 256

APPENDICES:

A. FREQUENCY DISTRIBUTIONS OF SAMPLE AND POPULATION DATA ON TWELVE SELECTED CHARACTERISTICS OF TEACHERS AND SCHOOLS 264

Page

APPENDICES:

B. LETTERS AND MEMORANDA 277

C. QUESTIONNAIRES USED 286

D. SUMMED INTERACTION MATRICES OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR THE YEAR FOR EACH OF SEVENTEEN REGENTS PHYSICS TEACHERS 298

E. FORMS AND MEMORANDA ASSOCIATED WITH CPT AND TOUS TESTING 316

LIST OF TABLES

Page

TABLE 1: STATISTICS ON SELECTED CHARACTERISTICS OF 131 PUBLIC HIGH SCHOOLS COMPOSING A 14-COUNTY CENTRAL NEW YORK STATE REGION 50

TABLE 2: COMPARISONS OF SAMPLE AND POPULATION DISTRIBUTIONS OF 12 PRE-SELECTED CHARACTERISTICS OF TEACHERS AND SCHOOLS USING THE KOLMOGOROV-SMIRNOV ONE-SAMPLE TEST 54

TABLE 3: STATISTICS ON SELECTED CHARACTERISTICS OF THE SAMPLE OF 17 SCHOOLS COMPARED TO THE POPULATION OF 131 SCHOOLS 56

TABLE 4: ADDITIONAL CHARACTERISTICS OF THE SAMPLE OF 17 PARTICIPATING HIGH SCHOOLS, 1964-1965 57

TABLE 5: SCORES ON TEACHER QUESTIONNAIRE 81

TABLE 6: STUDENT SCORES ON OBSERVER INTERFERENCE ITEMS. 83

TABLE 7: INTRA-OBSERVER RELIABILITIES ON RANDOMLY SELECTED MATRICES 90

TABLE 8: INTRA AND INTER-OBSERVER RELIABILITIES ON NONRANDOMLY SELECTED MATRICES 93

TABLE 9: DISTRIBUTION OF INTERACTION TIME AMONG MAJOR ACTIVITIES IN MINUTES 113

TABLE 10: GRAND INTERACTION MATRIX FOR ALL OBSERVATIONS OF 17 PHYSICS TEACHERS IN ALL MAJOR ACTIVITIES 117

TABLE 11: SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 15 PHYSICS TEACHERS IN ACTIVITY ONE - DEMONSTRATION 123

TABLE 12: SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 17 PHYSICS TEACHERS IN ACTIVITY TWO - LECTURE 124

TABLE 13: SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 15 PHYSICS TEACHERS IN ACTIVITY THREE - LABORATORY 125

TABLE 14: SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 16 PHYSICS TEACHERS IN ACTIVITY FOUR - RECITATION 126

Page

TABLE 15: CELL-WISE COMPARISONS OF GRAND MATRICES OF HIGH SCHOOL PHYSICS, EIGHTH GRADE MATHEMATICS AND SEVENTH GRADE SOCIAL STUDIES TEACHERS 132

TABLE 16: COMPARISONS OF SELECTED CELL AND COLUMN PERCENTAGES FROM COMPOSITE FLANDERS MATRICES.. 135

TABLE 17: COMPARISONS OF SELECTED I-D PERCENTAGES FROM COMPOSITE FLANDERS MATRICES 137

TABLE 18: DARWIN CHI-SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY ONE - DEMONSTRATION 143

TABLE 19: DARWIN CHI-SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY TWO - LECTURE 144

TABLE 20: DARWIN CHI-SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY THREE - LABORATORY 145

TABLE 21: DARWIN CHI-SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY FOUR - RECITATION. 146

TABLE 22: DARWIN CHI-SQUARE COMPARISONS OF SUMMED FALL MATRICES AND SUMMED SPRING MATRICES WITH ACTIVITY 3 (LABORATORY) EXCLUDED 149

TABLE 23: SCORES ON STUDENT QUESTIONNAIRE BY CLASSES ... 152

TABLE 24: CPT AND TOUS PUBLISHED RELIABILITIES 163

TABLE 25: CPT AND TOUS RELIABILITIES BASED ON PRE AND POST ADMINISTRATIONS 165

TABLE 26: APTITUDE AND ACHIEVEMENT CHARACTERISTICS OF REGENTS PHYSICS STUDENTS 169

TABLE 27: PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN REGENTS PHYSICS (JUNE 1965) SCORES AND STUDENT APTITUDE AND ACHIEVEMENT SCORES 172

TABLE 28: INTER-CORRELATION COEFFICIENTS OF 15 CONTROL VARIABLES AND TOUS-TOTAL PRE-TEST BASED ON 81 SETS OF OBSERVATIONS 176

TABLE 29: LINEAR MULTIPLE REGRESSION ANALYSES OF TOUS-TOTAL PRE-TEST ON VARIOUS COMBINATIONS OF CONTROL VARIABLES USING 81 OBSERVATIONS 184

Page

TABLE 30: LINEAR MULTIPLE REGRESSION ANALYSES OF TOUS-TOTAL PRE-TEST USING 203 OBSERVATIONS ... 186

TABLE 31: LINEAR MULTIPLE REGRESSION ANALYSES OF CPT PRE-TEST USING 208 OBSERVATIONS 188

TABLE 32: LINEAR MULTIPLE REGRESSION ANALYSES OF PHYSICS REGENTS (JUNE 1965) USING 178 OBSERVATIONS ... 189

TABLE 33: LINEAR MULTIPLE REGRESSION ANALYSES USED TO ADJUST TOUS AND CPT PRE-TEST AND PHYSICS REGENTS (JUNE 1965) SCORES WITH NYSRS-APT. (X_1), NYSRS-ACH. (X_2) AND REGENTS INDEX (X_3) AS INDEPENDENT VARIABLES 191

TABLE 34: CLASS MEAN ADJUSTED SCORES ON PHYSICS REGENTS (JUNE 1965) 193

TABLE 35: CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON COOPERATIVE PHYSICS TEST, FORM Z, (CPT) ... 195

TABLE 36: CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TEST ON UNDERSTANDING SCIENCE, FORM W, (TOUS) TOTAL SCORE 196

TABLE 37: CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TEST ON UNDERSTANDING SCIENCE, FORM W, (TOUS) PARTS I, II AND III 197

TABLE 38: GRAND MEANS OF ADJUSTED PHYSICS REGENTS SCORES AND OF ADJUSTED CPT, TOUS-I, II, III, AND TOUS-TOTAL PRE-POST GAINS 199

TABLE 39: INTER-CORRELATION COEFFICIENTS OF ADJUSTED PHYSICS REGENTS SCORES AND OF ADJUSTED CPT AND TOUS PRE-POST GAINS 200

TABLE 40: STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED SCORES ON PHYSICS REGENTS (JUNE 1965) 205

TABLE 41: STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON CPT 206

TABLE 42: STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TOUS-I, II, III AND TOTAL 207

Page

TABLE 43: RANKS ON CLASS MEAN ADJUSTED SCORES OF PHYSICS REGENTS AND OF CPT, TOUS-I AND TOUS-TOTAL PRE-POST GAINS 208

TABLE 44: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN TOTAL TEACHING 214

TABLE 45: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN LECTURE ACTIVITY 215

TABLE 46: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN LABORATORY ACTIVITY 216

TABLE 47: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN RECITATION ACTIVITY 217

TABLE 48: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSa 220

TABLE 49: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSb 221

TABLE 50: MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSc 223

TABLE 51: KENDALL'S TAU CORRELATION COEFFICIENTS OF TEACHER EFFECTIVENESS SCORES WITH MEASURES FROM FLANDERS MATRICES AND OTHER TEACHER AND SCHOOL FACTORS 225

TABLE 52: MANN-WHITNEY U TEST COMPARISONS OF TEACHER EFFECTIVENESS SCORES USING HIGH VERSUS LOW GROUPS ACCORDING TO SELECTED CORRELATED VARIABLES 230

LIST OF FIGURES

Page

Figure 1. Areas of the Flanders matrix................ 98

Figure 2. Distributions of tallies in Flanders categories....................... 127

Figure 3. Type of teacher statements as per cent of teacher talk.......................... 128

CHAPTER ONE

THE RESEARCH PROBLEM

Introduction

The conceptualization of instruction is one of the most significant tasks facing researchers in education today. Much progress has been made in the development of theories of learning; but learning and teaching are different. There is much to be done in the development of sound theories of teaching. In fact, only recently has it been possible to move forward with the initial phases of research, prerequisite to the construction of a theory of teaching, through developments of systematic classroom observation (Medley and Mitzel, 1958; Hughes, 1959; Smith, 1959; Flanders, 1960, 1964; Wright and Proctor, 1961; and Bellack and Davitz, 1963). With the availability of research tools for systematically observing and analyzing what takes place in the classroom, the problem of determining the nature of effective science instruction can be met head-on. This is an immediate task that must be undertaken to contribute a foundation for the subsequent development of a theory of teaching and it is one to which this research was addressed.

Scope of the Problem

Little is known with certainty about the nature of effective teacher behavior. Flanders (1963a) points out that "much of what is learned in education courses is neither conceptualized, quantified, nor taught in a fashion that builds a bridge between theory and practice" (p. 1). This may be harsh criticism, but there is no denying the lack of information relating teacher behavior to learning. Emphasizing the need to verify concepts in education by personal field experiences and expressing concern for the "increasing gorge between theory and practice," Flanders calls for a need to describe teaching as a series of acts through time and to "establish models of behavior which are appropriate to different kinds of teaching situations" (1963a, p. 1). Bellack and Davitz (1963) cite the need to go into the classroom and find out "who speaks, about what, how much, when, under what conditions, and with what effect." Medley and Mitzel (1963) add additional emphasis:

> Certainly there is no more obvious approach to research on teaching than direct observation of the behavior of teachers while they teach and pupils while they learn. Yet it is a rare study indeed that includes any formal observation at all. In a typical example of research on teaching, the research worker limits himself to the manipulation or study of antecedents and consequents of whatever happens in the classroom while the teaching itself is going on, but never once looks into the classroom to see how the teacher actually teaches or how the pupils actually learn (p. 247).

In another writing, Medley (1963) points out:

> Just about everything we know today (or think we know) about teaching and learning in the classroom is based either on analogous reasoning from research done outside the classroom or on somebody's opinion about what he saw in the classroom. When we are able to base our knowledge on objective observations, then and only then can we begin to build a science of teaching (p. 273).

There is no doubt about the need to study teaching--and certainly the place to study it is in the classroom where it exists. With the advent of reliable methods of systematic observation, objectivity can be realized and useful information can be collected and analyzed.

The concern of this research is high school physics teaching. Extensive innovations and revisions involving secondary physics curricula, facilities, and instructional aids have been and are continuing to be developed. But the store of reliable and useful information concerning teacher and student behavior in the physics classroom is relatively vacuous.

There is considerable agreement among science educators on particular factors necessary for effective science teaching. These are in the realm of such things as techniques of planning, grading, testing, demonstrating, assigning, classroom managing, etc. (Johnson, 1963a). We can include characteristics such as voice and diction, and perhaps, physical facilities, subject matter preparation and curricula. These things are important. But are they sufficient? How should the physics teacher interact with the students while engaging in the teaching process? Should he be more indirect? Should he be more flexible with his influence in the classroom? What is the nature of his influence and how is it related to achievement of accepted objectives? In short, can we learn more about teacher behavior--about what good teaching is?

This project was focused on an effort to determine what the physics teacher does in terms of verbal interaction while engaging in the communication of classroom discourse. A distinction between teacher intent and what actually takes place was planned. Reliable information about the types and patterns of verbal interaction which are characteristic of the more effective physics teacher was sought. The research tool of interaction analysis, as developed by Flanders, was used. Systematic observation of a number of physics teachers while they were teaching facilitated a sound attack on the problem.

Purpose of the Study

The purpose of this study was to identify, through systematic classroom observation, the characteristics and patterns of teacher behavior that exist in the high school physics classroom and to associate them with aspects of teacher effectiveness. The behavioral dimension involved was teacher influence through verbal interaction; the Flanders system of interaction analysis was employed. An effort was made to distinguish the more effective physics teacher from the less effective teacher and to seek relationships between teacher effectiveness and both "directness" and "flexibility" of teacher influence. In addition, it was planned to provide a large store of reliable data on teacher and student verbal behavior in the high school physics classroom which will serve as bases for further analyses at a future time.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Introduction

There is an extensive history of research efforts directed to the task of determining factors associated with effective teaching (e.g., Barr, 1935, 1961). The concern of researchers has embraced all areas of instruction including teaching at the college level (e.g., Wispe, 1951). Over the years numerous associated studies have focused on problems concerning objectives of teaching, curricula, methods, and facilities; others have dealt with characteristics of the teacher and student. A large portion of the reported results is based on opinions and judgments. In much of the remainder statistical techniques of various degrees of sophistication have been employed to manipulate and control selected variables in order to acquire reliable information concerning teaching effectiveness. The extensive work of Hurd (1930, 1933, 1934) is illustrative of the early efforts in the area of physics teaching; efforts such as those of Ort (1964) and of Mallinson (1965) may be cited as representatives of current attacks on the problem. No single factor has yet emerged as a necessary and sufficient condition of teacher effectiveness. It is not likely that one will.

"How can we best educate our young people?" is a question for which there is no single or final answer. It is one that has been asked over a long time; hopefully, it is an inquiry we will continue to make. The continual acquisition of many partial answers facilitates progress toward better education of our young people.

Currently, the most promising area of research on teaching appears to be the behavior of teacher-pupil interaction in the classroom. Various behavioral dimensions, from both the affective and substantive standpoints, merit study. Few of the earlier efforts have brought investigators into the classroom situation where the teaching process exists. The implantation of the seeds of current efforts and possibilities is relatively recent (e.g., Hare, Borgatta, and Bales, 1955). Social interaction in the classroom, well summarized by Withall and Lewis (1963), has become a front of promising research possibilities through the development of techniques of systematic observation of teacher and pupil behavior.

The Need for Systematic Observation

Attempts to measure teacher effectiveness through the employment of criteria based on pre-conceived notions of effective teaching have produced little reliable data. The employment of rating scales, wherein a teacher is observed and then judged, has not provided information that reliably distinguishes between the behavior of effective and ineffective teachers. This does not imply that all rating scales are valueless; on the contrary they can be quite useful if employed properly in selected situations for particular guiding purposes. What is placed in question is their usefulness for the study of teacher behavior. Ratings of effectiveness compared with criteria of effectiveness based on aspects of pupils' learnings have given, as Medley and Mitzel point out, "uniformly negative results" (1963, p. 257). The point is made that even though measurements of student achievement are imperfect criteria "in the sense of measuring growth toward all objectives of instruction," they should show "at least moderate" correlations with judgments of teacher effectiveness. They do not.

A major attempt to identify behavior patterns discriminating effective from ineffective social studies teachers was carried out by Barr (1929). Although the criterion of teacher effectiveness was based on supervisory judgment and not on student achievement, the study "marshaled what was probably the greatest variety of behavior data ever brought to bear on the problem of distinguishing these two types of teachers" (Medley and Mitzel, 1963, p. 258).

In developing their plea for systematic classroom observation wherein teacher behavior is objectively observed in short periods of time and critically analyzed later, Medley and Mitzel are emphatic:

> The fact that ratings of teacher effectiveness have no discernible relationship to effectiveness does not mean that effectiveness cannot be measured in process. Since

> it may be assumed that whatever effect a teacher has on pupils must result from his behaviors, it is only necessary to identify the crucial behaviors, record them, and score them properly to measure effectiveness in process (1963, p. 258).

Studies of Teacher Behavior Employing Systematic Observation

The emphasis on objectivity of observation was initiated by Thomas, et al. (1929) in their study of child behavior in the nursery school. The purpose was to study pupil-pupil and pupil-teacher interaction. Having rejected descriptive accounts and diaries of behavior, they sought to obtain data on social behavior which would be independent of observers within a small and predictable range of error. Various schemes were employed to obtain a high observer agreement but ambiguity of definition (for example when a child does and does not make a social contact) was found to cause low reliability. They noted that "these low coefficients of correlation are a beautiful example of how unreliability creeps in where interpretation is permitted the author" (1929, p. 9).

Anderson and Brewer (1945), in extending research on nursery school children, developed a scheme of systematic observation employing periodic categorizations of behavior to distinguish pupil-teacher contacts on a dominant-integrative behavioral dimension. They also developed a method for simultaneously observing pupil and teacher behavior to detect effects of teacher behavior on pupils. One child at a time was observed for a five-minute period and records of behavior and teacher contacts were produced. A total of two hours of observation, in five-minute segments, was made on each child. Reliability associated with observer agreement on the categorizations made (item by item) was high. In addition, the ratio of integrative contacts to dominative contacts (frequency) led to a so-called "I-D Index" which served as a variable in comparing teachers and classes. They were able to show differences between classes in quantitative terms and related their "I-D Index" to teacher personality. Although the classes of only four teachers were studied, the development of the notion of objective observation through the coding of behavior, followed by analysis, and the effort to quantify behavioral measurements were major contributions to early research on teaching.

Withall (1949), referring to the dominative-integrative behavioral dimension as "Social-Emotional Climate," developed a categorization scheme to study teacher behavior only. It contained seven categories of behavior extending from extremes of learner-centeredness to those of teacher-centeredness and was the basis of a so-called "Climate Index." The categorization was used to code typewritten transcripts of classroom discourse. His concern for reliability in terms of consistency of teacher behavior from one sample of behavior to another (as opposed to reliability limited to observer agreement on categorizations of a particular sample) was an important contribution. He used his instrument (1951) to study the classroom climates of different seventh-grade teachers teaching the same group of pupils by following a group from subject to subject. His results showed that for a particular group there was a variation of climate from teacher to teacher.

Medley and Mitzel (1958) were particularly concerned with the reliability of Withall's categorization scheme. It is to be emphasized that reliability in this sense is distinguished from observer agreement. A high reliability would indicate consistency of measured behavior. The reliability coefficients of concern to Medley and Mitzel "estimated the correlation between the number of statements classified in a particular category by one observer on one visit on the basis of approximately 100 consecutive statements with the number classified in the same category by a different observer on a different visit" (1963, p. 268). Low reliabilities were found and it was felt that this reflected "instability of behavior rather than any difficulty experienced by the observer in using the technique" (1963, p. 268). It was concluded that two observers categorizing the same behavior could provide little more information than one, and that it was better to observe the teacher more often.

They showed that while reliability increased with the number of visits, the rate of increase leveled off after about twelve visits. Obtained reliabilities in this case extended from 0.75 to 0.90.

Later, using modifications of items developed by Cornell, Lindvall, and Saupe (1952), and guided by experimentation with Withall's technique, Medley and Mitzel (1959) perfected their instrument for systematic observation (Observation Schedule and Record, or OScAR). The instrument was designed to provide "quantitative data regarding behaviors of beginning teachers so that the behaviors could be correlated with a number of other variables" (1963, p. 278). Designing the instrument for use by a single observer working alone, they simplified category definitions and separated the process of scoring from the process of observing. No behavioral dimension or scale was initially involved.

The instrument was used to study forty-nine beginning elementary teachers in nineteen New York City schools over a period of ten weeks. Each teacher was visited twice by each of six observers working individually. Combining the OScAR items into "keys" (on *a priori* judgments that they belonged together) fourteen keys were ultimately produced. Mean key scores showed reliable differences between classes.

By means of a centroid factor analysis, the fourteen keys ultimately led to three scales: "Emotional Climate," "Verbal Emphasis," and "Social Structure." The first is associated with the amount of hostility observable in the classroom, while the latter two refer to the degree to which verbal activities predominate (pupil reads, writes, or textbook is used, etc.) and, respectively, to the grouping of pupils. The results showed "Emotional Climate" to be most reliable (0.90). Unfortunately, the three scales were not related to measures of pupil achievement. However, they did account for some of the teacher differences associated with student rapport and supervisors' judgments of teacher effectiveness.

The OScAR technique is an interesting one. During a five-minute interval of observation, the following is checked as observed: type of activity (extending from "teacher works with individual pupil," to "teacher demonstrates," to "pupil laughs," etc.); type of grouping; instructional materials being used; and "signs" (such as "teacher shows affection for pupil" and "pupil shows hostility for teacher"). No item is checked more than once in the five-minute period. As soon as the five minutes are up, the subject (math, social studies, etc.) is checked. For the next five minutes, the observer tallies teacher statements according to the categories of "Pupil-Supportive," "Problem-Structuring," "Miscellaneous," "Directive," and "Reproving." Simultaneously, non-verbal indicators of teacher approval or affection for the student (looks, gestures, etc.) are tallied under a single category. Similarly, frequencies of indicators of hostile or reproving behavior and teacher authority are recorded under separate categories. At the completion of the second five minutes the entire process is repeated a second and third time.

In her study of student teacher behavior in high school classes, and employing a form of OScAR, Morrison (1961) confirmed Medley and Mitzel's basic conclusions. Although OScAR has distinct limitations, it does represent a good beginning.

A "Model of Good Teaching" was produced by Hughes and her associates (1958, 1959) in their study of elementary teachers. Differences in the patterns of teacher acts (functions) were sought among teachers who were "judged good" and among teaching situations. In particular, the question was, "What is good teaching?"

A total of thirty-one "judged good" teachers (six in a pilot study) were studied to develop the model. In addition, five pairs of teachers of grades 1-5 in a single school (a "best" teacher and one of "most concern") were compared. Three thirty-minute records were acquired on each teacher--there was one each of a reading lesson, a developmental lesson in social studies or arithmetic, and an activity period. For each record two observers simultaneously recorded a handwritten account using a shorthand notation. No pair of observers visited the same teacher twice.

Observed functions were identified according to the "University Revision of the Provo Code for the Analysis of Teaching." This code consists of seven function categories involving a total of thirty-three specific functions which teachers perform in the classroom in interaction with children. These are: "Controlling Functions," "Imposition of Teacher," "Facilitating Functions," "Functions that Develop Content," "Functions that Serve as Response," "Functions of Positive Affectivity," and "Functions of Negative Affectivity." A total of 129 records were analyzed and a "Dominative Behavior Index" was used.

Some of the major findings were as follows:

1. In nearly 2/3 of 129 records, control functions which limit the extent to which children can "explore ideas, reach out in their experience, and work on their own" exceeded 40 per cent of all teaching acts.
2. Functions that develop content were infrequently used (less than 20 per cent of all acts in 3/4 of the records). A lot of teaching for rote learning is implied.
3. Teachers seldom responded to pupils and half the time functions of evaluation were not accompanied with reasons for correctness or incorrectness. Also, teachers were more positive than negative (more supportive in affectivity).
4. In 80 per cent of all records teachers were dominant in over 50 per cent of their total teaching acts; 3/4 of the teachers had 40 per cent or less of their total teaching acts classified as integrative.

5. Primary teachers were significantly more controlling and more negative than intermediate teachers.

Hughes pointed out that "one of the major skills to be acquired is that of responsiveness to children in a discriminating professional manner." None of the seven functions discriminated between the group of twenty-five "judged good" teachers and the ten "representative" ones.

Hughes' construction of a model of good teaching behavior is particularly interesting. The conclusion was that the following percentage ranges, in each category, produced the best pattern of interaction: Controlling, 20-40 per cent; Imposition, 1-3 per cent; Facilitating, 5-15 per cent; Content Development, 20-40 per cent; Personal Response, 8-20 per cent; Positive Affectivity, 10-20 per cent; and Negative Affectivity, 3-10 per cent.

It is likely that at best these "conclusions" can be viewed as hypotheses; however, they are nevertheless challenging. A strong point is made by Hughes that conclusions pertaining to elementary school teaching would apply to all levels--including graduate school! Although debatable, it is certainly something to think about.

Smith's study (1959, 1963a, 1963b) of the logic of classroom discourse represents an attack on the problem from the standpoint of behavioral units. By tape recording the discourse of seventeen Illinois high school teachers (grades

9-12 and across English, mathematics, science, and social studies) for five consecutive class sessions each, and by supplementing the resulting written transcripts with the running commentary of a classroom observer, Smith was able to analyze the ensuing verbal behavior in terms of "episodes" and "monologues." The episode is defined as a verbal exchange between two or more speakers while the monologue is an individual's contribution, or "solo performance." For example, the extended lecture would be an extensive monologue. A five-year project, Smith's major purpose was to "identify and describe the variables which make up teacher behavior" (1963, p. 1). "Strategies," or large scale maneuvers of logical operations, i.e., patterns of "acts that serve to attain certain outcomes and guard against others" (1963, p. 4) were a basis for analysis.

In comparing the "closed episode" (viz., a question eliciting a single response) to the "open episode" (such as an open-ended question which will stimulate discussion), Smith speculates that openness of episodes encourages creativity in the various sciences. His episodes are analyzed through the categorization of "logical operations" such as defining, describing, designating, stating, reporting, comparing and contrasting, opining, explaining, etc. He states that "the identification of units of behavior within which these elements (logical operations) have meaning is one of the main tasks in the conceptualization of instruction" (1963b, p. 297).

Drawing in part on the work of Smith, Bellack and Davitz (1963) studied a group of fifteen high school social studies teachers who taught a specially prepared four-day unit on economics. Their objectives were to:

> . . . describe the verbal events that occurred in each class, to discover similarities and consistencies in the pedagogical patterns of discourse, to define the distinctive aspects of the roles played by teachers and pupils, and to provide some estimate of the variability among teachers and classes along each of the dimensions of discourse investigated (1963, p. 128).

Taking a "game theory" approach to teaching, they analyzed transcripts of the discourse from the standpoint of pedagogical "moves" of structuring, soliciting, responding, and reacting. Cyclical patterns or combinations of pedagogical moves, called "teaching cycles" were identified. Quantification was realized in terms of lines of discourse. Teacher effectiveness was measured by pre-post achievement gains (students were tested on the unit immediately before and after the four days of instruction) and with a subject matter attitude scale. Scores were adjusted on the bases of intelligence and community socio-economic status.

A principal finding was that there was a "striking intrateacher consistency" of "Emotional Meaning," a dimension somewhat comparable to Medley and Mitzel's "Emotional Climate." Although able to relate few aspects of behavior to student learning, the technique of discourse analysis they developed (their prime objective) is itself a major achievement.

They found that even though teachers were given a specific outline of material to treat, there were "marked differences in the substantive material covered in the class sessions" (1963, p. 104). This variation of content taught likely impeded the ability of the content test to distinguish classes on the basis of learning. A major point of emphasis, however, is the increasing promise of systematic observation. As Bellack and Davitz put it:

> Indeed, at the current stage of knowledge about the teaching process, a variety of widely different approaches to the same problem is probably desirable. Nevertheless, experience in this study would suggest that studies within actual classroom settings, with some degree of control, is an especially useful mode of investigation in this area of research (1963, p. 167).

Thus the work of Smith suggests the need to study teacher behavior in particular subject matter areas and that of Bellack and Davitz yields one method of doing it.

Attention was given to secondary (and college) mathematics by Wright and Proctor (1961) who have developed a complex scheme for systematically observing mathematics teaching. Their technique involves consecutive observation and recording in a series of adjacent fifteen-second intervals throughout the class period (extensive categorization). The major dimensions employed are "Control," "Process," and "Attitude"; scores are counts and proportions in each category (numerous sub-categories are involved). Categories are not mutually exclusive.

Twelve mathematics classes in St. Louis were analyzed with "degree of mathematical rigor" and "amount of pupil participation" (low and high in both) as independent variables. The results showed that groups could be discriminated and differences in various styles of teaching could be clarified in mathematics classes.

Aschner and Gallagher (1962), basing a categorization system on Guilford's model of the "Structure of Intellect" (1960), followed Smith's technique in studying transcribed discourse of secondary school teaching. Later, Aschner (1963) used the system in selected junior high school classes to study the relationship between student initiative and intelligence. Initial findings have confirmed hypotheses of positive relationships. The system clearly distinguishes teacher behavior which restricts student participation from that which expands it.

Much of the foregoing (Anderson, Withall, Medley and Mitzel, Hughes, Bellack, Aschner) suggests directness of teacher influence (integrative versus dominative, positive affectivity versus negative, emotional meaning, emotional climate, teacher centeredness versus learner centeredness, etc.) as a dimension of teacher behavior that may be related to achievement in high school physics. Both analysis of verbal interaction and analysis of discourse logic are promising approaches to the task of learning more about the nature of effective (and ineffective) teaching. Perhaps the work of Aschner and of Wright lies somewhat in the middle.

Each of the studies was uniquely productive--all of them capitalized on some method of systematic classroom observation. Little use has been made of any of the techniques in the secondary science area. The most promising for the purposes of this study is the method of Flanders (1964)--but first some comments on a few pertinent studies not employing systematic observation.

Studies on Science Teaching

Research on secondary science teaching is largely limited to relatively futile attempts to relate various aspects of the teacher and the classroom to effectiveness. Effectiveness has generally been assessed through judgment based on preconceived notions of effective teaching. Pupil gain scores on subject matter achievement tests, also a limited criterion, are commonly used.

Citing the scarcity of research on science teaching, Watson points out that:

> Nearly all individuals in a position to conduct research on teaching science were originally committed to natural science and its teaching, rather than to the behavioral sciences. Obliged to maintain some comprehension of explosively growing scientific knowledge, and involved in teacher training and retraining through courses and numerous institutes, such individuals have had little time or inclination to become competent in the aspects of behavioral science research needed to study the consequences of various forms of teacher behavior. To acknowledge this is not to find fault with science educators, but rather to make clear why contributions to research on teaching science have been so meager. Forthright acknowledgement of past omissions may spur increased research activity in this field (1963, p. 1032).

Research focusing on high school physics teaching is particularly limited.

An approach avoiding some of the pitfalls of observer judgment of teacher effectiveness has been the use of students' perceptions of teacher behavior. The rationale that the pupil's image of the teacher is the important image and that pupils observe the teacher extensively has some merit. It also has the "catch" that the reported teacher behavior is second hand.

Cogan (1956, 1958, 1963) studied thirty-three teachers, four of whom were science teachers, and their 987 eighth-grade students (two socio-economically different New England communities). He found that pupil reports of "inclusive" teacher behavior (indirect or integrative behaviors tending to expand pupil participation in the classroom process) were positively correlated with the amount of pupils' reported self-initiated school work wherein the latter is assumed to be a measure of teacher effectiveness. The same conclusions held for the sub-sample of four science teachers.

Reed (1961) studied thirty-eight ninth-grade general science classes (38 teachers, 19 schools, and over 1,000 students). Student interest in science was used as a measurable objective of science teaching. Three teacher behavioral criteria were employed: "warmth," "demand," and "creation of intrinsic motivation." All were reported as perceived by the students. A seventy-item "Science Activity Inventory," developed by Reed, assessed pupil interest in science by respondents' indications of various science related activities undertaken during the school year. Pupils' perceptions of the independent variables were obtained from an attached check list. Significant within-class correlations were found for "warmth" and "interest" (boys, 0.20; girls, 0.28). "Demand" was unrelated to "interest"; "intrinsic motivation" was positively related.

Both studies may be plagued by a tendency for students to make socially desirable responses. This is an inherent limitation in getting at teacher behavior through the "pupils' eyes." But possible support of the teacher influence dimension is of particular interest.

The broad analysis of teacher characteristics made by Ryans (1960, 1963) included some 500 secondary mathematics and science teachers. Three dimensions of teacher behavior were painstakingly developed: "warm, understanding, friendly behavior"; "responsible, businesslike, systematic behavior"; and "stimulating, imaginative surgent behavior." An approach was developed to carefully associate the dimensions to effectiveness; limitations were recognized.

Trained observers rated teachers on the dimensions. Teachers rating very high on all three scales were compared to those uniformly rated as very low. Detailed analyses of extensive arrays of characteristics of all secondary teachers (across all subject levels) led to a set of distinctly consistent traits for the "high" group and, similarly, for the "low" group.

Ryans reported that "high" group members more frequently "manifest extreme generosity in appraisals of the behavior and motives of other persons; express friendly feeling for others; indicate preference for student-centered learning situations; manifest independence, though not aggressiveness; indicate greater preference for non-directive classroom procedures; indicate greater enjoyment of pupil relationships" (1960, p. 360). This represents almost 50 per cent of the characteristics reported.

Characteristics of the "low" group were to the contrary, i.e., "low" group members more frequently "are restricted and critical in appraisals of the behavior and motives of other persons; indicate preference for teacher-directed learning situations; indicate preferences for activities which do not involve close contacts with people" (1960, p. 360).

The striking implication is that secondary teachers who rated "high" (upper 2%) on the criteria seem to have characteristics that can be associated with indirect teacher influence while, conversely, those rated "low" (lower 2%) were direct. It seems reasonable to assume that secondary science teachers in New York State will be similar to Ryans' group.

The Work of Flanders

One can only speculate about directness of influence in the case of physics teaching. The implication that it is an important dimension to investigate in the physics classroom is, to this author, a common "thread" throughout the research. The need to study the dimension through systematic classroom observation is also clear. Flanders' method of interaction analysis (1964), a distinct "breakthrough" as far as this writer is concerned, provides the means. Interestingly enough, it has yielded information to indicate that flexibility of influence may be more important than directness.

Devised and proven through eight years of development, the Flanders technique (Amidon and Flanders, 1963; Flanders, 1964a) provides a means of critically analyzing verbal interaction in the classroom. Verbal interaction, assumed to be an adequate sample of total classroom behavior, can thus be readily studied, analyzed and assessed.

In facilitating the study of teacher behavior on the dimension of directness of teacher influence, Flanders provides ten mutually exclusive categories which are intended to be inclusive of all teacher-student verbal behavior. Four categories associated with teacher statements that encourage or expand student participation (indirect teacher influence) are: "Accepts Feeling," "Praises or Encourages," "Accepts or Uses Ideas of Students," and "Asks Questions."

The remainder of teacher talk is classified according to: "Lecturing," "Giving Directions," and "Criticizing or Justifying Authority." The latter three correspond to teacher statements that restrict student participation and consequently are associated with direct teacher influence. Two categories distinguish student talk: "Student Talk--Response," and "Student Talk--Initiation." A tenth category, "Silence or Confusion," is associated with short periods of silence and short periods of confusion in which "communication cannot be understood by the observer" (1964a, p. 5) within sessions of verbal interaction.

Behavior is categorized every three seconds; the observer records the category number of one three-second interval while observing the next interval. Consecutive numbers are plotted into a 10 X 10 matrix. For example, the recorded sequence 4, 8, 2, 5, 4 would yield one tally each in matrix cells 4-8, 8-2, 2-5, and 5-4 (x-y means row x and column y). The sequence is that a teacher asked a question, a predictable student response followed, the teacher countered with a praise statement followed by a lecture statement (teacher's own idea, fact, or opinion) and then the teacher asked another question.

Since a time scale is employed, the examination of tallies of columns, column and row segments, and individual cells can provide measurements of the relative extent to which various aspects of teacher-student verbal interaction take place. For example, one can determine the percent of the time that the teacher lectures, asks questions, responds to predictable student statements (coming from narrow questions) with praise statements, uses extended direct influence, uses extended indirect influence, etc.

It is important to emphasize that the Flanders method gives the amount, sequence, and pattern of verbal behavior along a dimension of directness (or indirectness) of teacher influence. Medley and Mitzel (1963) cite it as "the most sophisticated technique for observing climate thus far. . . ." Their review is pertinent:

> Flanders' scheme is extremely ingenious. Every one of the 100 cells in the matrix . . . represents a different item of behavior with its own intrinsic interest. Yet the observer needs to learn and use only 10 categories. The idea of categorizing the dominant pattern of a three-second period rather than each statement or other unit of behavior is also ingenious (1963, p. 274).

Of additional interest is Flanders' view that any appropriate set of categories could be used in this fashion. The studies of Bales (1950), Anderson, Withall, and Mitzel have contributed to his system.

It suffices to report briefly some recent research results in which Flanders' method has been employed. It is of interest to preface the remarks by noting that in an early laboratory study (1951), Flanders investigated the effects of domination on individual students. From galvanic skin measurements and heartbeat rate indications he was able to show that sustained patterns of domination were disliked, produced anxiety, and reduced recall.

In a study of 140 eighth-grade students who had scored highly on a specially prepared test of dependent-proneness, Amidon and Flanders (1961) employed independent variables of teacher influence (direct and indirect) and goal perception (clear and unclear); the dependent variable was content achievement in a unit on geometry. Teacher influence was measured and controlled through Flanders' method of interaction analysis. Instruction for all groups was carried out in a two-hour sequence of treatment and testing.

Results showed that on the post-achievement test the mean score of the indirect influence group was significantly higher than that of the direct influence group (F = 7.76, significant at the 0.01 level). The difference was significant at the same level when pre-achievement and IQ scores were controlled by analysis of covariance. It was concluded that "dependent-prone students learned more in the classroom in which the teacher gave fewer directions, less criticism, less lecturing, more praise, and asked more questions which increased their verbal participation" (1961, p. 4).

The foregoing is part of a two-year study of teacher influence directed by Flanders (1960, 1964) in Minnesota and New Zealand. In a field experiment approximately 450 seventh-grade social studies and as many eighth-grade mathematics students were involved as were their thirty-two teachers. Classes were observed during the first two days, middle two days and last two days of a two-week unit (sixteen mathematics classes and sixteen social studies classes). Teacher-student verbal interaction was recorded with the Flanders system. Since no interaction data can be taken during prolonged periods of silence (e.g., students doing seat work), the six periods provided about three hours of interaction data on each teacher.

It was found that the more indirect teachers were most indirect while clarifying content goals and while introducing new material. They became much more direct when work was in progress. Also, their students had significantly higher subject matter achievement--this held for both mathematics and social studies. Achievement differences favoring the more indirect teachers were "beyond the 0.01 level of confidence" (1964, p. 12). In addition, pupils of the more indirect teachers had significantly higher scores (more desirable attitudes) on an attitude inventory (better than 0.01 level).

Of particular interest is Flanders' finding that the concept of teacher flexibility of influence may be more important than directness of influence as a factor of achievement. He found the more successful teachers to be more flexible, i.e., they could be as direct as any teacher in certain situations but were more indirect in other situations. Concern for different "situations" is motivated, for example, by his finding that higher achievement and less dependence were associated with goal clarification through indirect influence (1963, p. 58). Flanders' own interpretations are pertinent:

> There seemed to be four essential elements of teacher influence in the classrooms in which achievement and attitudes were superior. First, the teacher was capable of providing a range of roles spontaneously in the classroom that varied from fairly active, dominative supervision, on the one hand, to reflective, discriminating support on the other hand. The teacher was not only able to achieve compliance, he was also able to support and encourage student initiative. Second, the teacher was able to control his own spontaneous behavior so that he could provide one role or another at will. Third, he must have had sufficient understanding of principles of teacher influence so that he could make a logical bridge between his diagnosis of the present situation and the alternative actions that he could take. He had a science to his art. And fourth, he must have been a sensitive, objective observer who could make valid diagnoses of current conditions. All of these attributes, among our most successful teachers, were super-imposed upon a firm grasp of the subject matter being taught.
>
> Teachers in our studies who failed to approach these requirements appeared to be restricted to a limited number of roles which produced a more rigid sequence of actions, in turn, there was little variation from one classroom situation to the next. In particular, the ability to expand the freedom of action of the students by self-control of verbal influence seemed missing in our less successful teachers (1964b, p. 19).

It might be added that it is Flanders' contention that the more indirect approach, in stimulating verbal participation by the students, is a means of providing the teacher with the students' perception of the situation (feedback). This gives the teacher more information and develops student responsibility. The more direct approach, on the other hand, increases student compliance and conditions the student to be more dependent. In addition, the indirect teachers are able to effectively use social skills associated with accepting, clarifying, and employing ideas and feelings of the students; as a consequence, they have less need for directing and criticizing.

A point emphasized is that there is no such thing as a purely direct or purely indirect teacher and that, although a teacher may be very direct or indirect over a short time period, every teacher blends direct and indirect influence into a "balance" which is stable over long periods of time. Although there may be less variability of the "balance" from teacher to teacher, teachers can be distinguished on the basis of net directness (or indirectness).

These findings are indeed exciting and challenging. They are, of course, based on limited data--much more needs to be done. The method has been especially useful in providing objective data. We need much more of it.

Some Theoretical Constructs of Teaching

Although the current status of teaching theory is one of infancy, it is intriguing to view Thelen's notions (1963) in the perspective of the empirical evidence found thus far. Thelen holds that teaching is "the facilitative reorganization of thoughts and feeling, and its essence is the control and direction of consciousness in planned or selected experiences." His teaching model involves two main types of activities: (1) discovery, and (2) dialogue which reflects discovered knowledge against established knowledge. "In one, the individual (student) is making discoveries. In the other, his discoveries and those of other individuals are confronted by established knowledge in such a way that personal knowledge is assessed, corrected, extended . . . new tensions generate new questions for investigation" (Thelen, 1963). According to Thelen, the teacher should establish a situation in which students freely discover (viz., a silent demonstration). This invites speculation. The teacher causes students to have "course-relevant ideas of their own in mind." A dialogue follows in which a wide range of testimony is invited; the teacher invites and guides students to examine their own testimony. It is in this phase, especially, that the teacher's understanding of the subject matter comes to bear in guiding the students by analyzing their responses. Identification and delimitation of major problems follow and "problem solving through discovery" ensues. Students then organize their awareness into communicable testimony and a second dialogue is conducted to "cast the students' discoveries in the problem-solving experience against the established discipline and methods of working" (Thelen, 1963). The model shows continual alteration of discovery and dialogue. The process must be infiltrated, according to Thelen, with meditation to avoid being just another "prescription of a set of procedures."

Thelen's model is an interesting framework upon which to speculate. One might just as well, perhaps, tend towards

Guilford's (1960) notions of convergent and divergent thought. One can detect a sense of accord between research findings that are currently materializing and the notions of the theorists. Through research on teacher behavior at all levels, progress toward theoretical constructs of effective behavior will be enhanced. In particular, extensive reliable and objective information on secondary science teacher behavior along promising dimensions is necessary to confirm or challenge the findings of others, and to provide useful implications for theory and practice.

Concluding Remarks

Some current research directions pertinent to secondary science teaching lie in both behavioral and non-behavioral areas. Perhaps the more pronounced changes that have materialized are the extensive curricula modifications that have come about, at all levels, with the advent of Physical Science Study Committee (PSSC) Physics. Although there is yet much disagreement on the directions that secondary science curricula should take, as evidenced by the report of the 1963 Ad Hoc Advisory Conference on Secondary School Physics (see National Science Foundation, 1963) for example, the extended interest of university (and other) scientists in the secondary science program is an asset to future progress. In the main, new programs are superior to old. The question of what to teach is receiving much attention (e.g., Brakken, 1965; Day, 1965).

But teaching is not telling. It is necessary to provide young people with well equipped classrooms and laboratories and with teachers well versed in subject matter who undertake instruction in accordance with improved science programs designed to meet sound objectives. However, it is not sufficient. Studies of teacher behavior are called for and the outlook of real progress from this standpoint is promising. It is encouraging to see accelerated interest in research connected with the training and supervision of teachers (e.g., Amidon and Simon, 1965; Furst, 1965; Hough and Amidon, 1965; Matthews, 1965; Zahn, 1965; and Blumberg and Amidon, 1965). In addition, numerous approaches to getting at what does happen in the classroom (e.g., Amidon, 1963; Parakh, 1965) may ultimately lead to valuable things that can happen there.

The long-range goal of research on teacher behavior is the development of sound theories of teaching. In a word, we are now able to begin to perform reliable and objective observations of aspects of teaching phenomena. In time, the collective "bits" of information, fruit of the efforts of many, will facilitate a realization of more effective science teaching.

Flanders' tool of interaction analysis makes an early step in a behavioral study of the physics phase of secondary science teaching possible. It is this setting that motivated the analytical research to be reported in what follows.

CHAPTER THREE

THE EXPERIMENTAL DESIGN

The purpose of this chapter is to provide a broad overview of the project's experimental design. Details will be supplied with data and results in succeeding chapters.

Objectives

This research project was essentially an analytical study. Its focus was on high school physics teaching. The objectives were:

1. To describe the characteristics and patterns of teacher-student verbal interaction in high school physics classrooms using the Flanders method of interaction analysis.
2. To determine relationships between teacher effectiveness, as measured by selected aspects of student achievement, and directness of teacher influence for total instruction and during selected major class activities.
3. To determine relationships between teacher effectiveness, as measured by selected aspects of student achievement, and flexibility of teacher influence.
4. To interpret all of the foregoing, and, by gaining insight into the nature of effective teacher behavior,

to produce implications for:

a. Supporting or challenging the findings of others.

b. Ways in which the education and supervision of science teachers can be improved.

c. Further research on teaching secondary science.

5. To provide a store of useful data on the discourse of high school physics classes which will facilitate further analyses at a future time.

Procedures

General Design

A sample of seventeen high school physics teachers of the New York State Regents Physics Course and their students were studied. Seventeen central New York State high schools were involved. It was assumed that: (1) the New York State Regents Physics Course is typical of high school physics curricula, and (2) verbal behavior in the classroom is an adequate sample of total teacher behavior.

Classes were systematically observed for four consecutive days in the fall and for three consecutive days in the spring. The first day served to acclimate the teacher and students to the observer's presence; a minimum of six class periods of observation provided the data. The observations focused on the teacher and the teacher-student verbal interaction that transpired in the classroom. All observations and subsequent analyses were done by this investigator.

In each case the discourse was recorded on magnetic tape and notations in the form of a running commentary were made in the classroom. The running commentary played a supplementary role in the investigator's laboratory in the course of a careful examination of the tape recordings. Verbal interaction was categorized according to Flanders' categorization which is based on directness of teacher influence, a behavioral dimension. Resulting Flanders matrices were combined in various useful ways, and, subsequently, analyzed in detail. Generalizations to the population, made possible through the sampling technique, were carried out. A grand composite matrix for all teachers in all activities (over 86,000 tallies) was compared to Flanders' composites for eighth grade mathematics and seventh grade social studies teachers.

An assessment of student attitude, using a specially adapted form of the Michigan Student Questionnaire, enhanced the interaction analyses.

In addition to serving the descriptive role of Objective One, grand matrices for each teacher from the fall observations were compared to spring composites. Darwin's likelihood ratio criterion (Darwin, 1959) was used to test the hypothesis of comparable frequency distributions of tallies. The extent to which such matrices were similar provided an indication of how well the distributions of tallies could be associated with a consistency of teacher behavior. In effect, it indicated how well the numbers recorded provided a picture of the teacher's style of teaching.

Teacher effectiveness was assessed in terms of the following aspects of class achievement: (1) understanding of facts and principles of physics as specified in the _Regents Physics Syllabus_ and as measured by the _Regents Physics Examination, June 1965_; (2) gain in understanding of facts and principles of physics commonly associated with high school physics courses as measured by pre-post gain scores on _Cooperative Physics Test, Form Z_; and (3) gain in understanding of the nature of science as measured by pre-post gain scores on _Test On Understanding Science, Form W_. In all cases class means of adjusted scores served as teacher effectiveness scores. Extensive aptitude and previous achievement data, obtained from permanent records in the schools, were used in a series of linear multiple regression analyses to adjust scores for student ability and previous performance in secondary mathematics and science courses. In cases of pre-post testing, the pre-test score was adjusted and the latter subtracted from the post-test score to provide an adjusted gain score. Pre-testing was done in October 1964 and post-testing was done in May 1965.

Both indirect-direct ratios (ratio of tallies from selected matrix cells associated with indirect teacher influence, denoted as "I," to those of direct teacher influence, "D") and indirect-direct percentages, i.e., I/(I+D) were calculated. However, only the latter were ultimately used in the analyses (the former may be convenient for future reference).

For each teacher a total I/(I+D) score (matrix column totals: columns 1..4 divided by columns 1..7) was obtained from the teacher's grand matrix of "total teaching" (based on at least six periods of observation). This ordinal-scaled score was employed as the measure of directness of teacher influence. The Mann-Whitney U Test (Siegel, 1956, p. 116) was used to test the null hypothesis of no difference of achievement between classes taught by "indirect teachers" and those taught by "direct teachers." The test was separately applied for directness based on each of three "major activities" as well as "total instruction" and for each aspect of achievement. The level of significance used for rejection in all tests in the study was 0.05.

In a similar fashion the same test was used to investigate flexibility of influence and achievement. In this connection three measures of flexibility of influence were explored.

For each aspect of achievement, Kendall Rank Correlation Coefficient Tau (Siegel, 1956, p. 213) was used to determine correlations between various I/(I+D) scores derived from Flanders matrices and adjusted class achievement scores. Numerous other correlations were determined as well.

Population and Sample

The population was defined as a central New York State region's 131 public high schools in fourteen counties surrounding Ithaca, New York. The sample of seventeen teachers and the corresponding seventeen high schools was constructed in such a way as to reflect, as a group, twelve selected characteristics of teachers and schools in the population. The extent to which the sample was statistically "as good" as a random sample was determined by a comparison of the sample to the population on each of twelve characteristics. The Kolmogorov-Smirnov One-Sample Test (Siegel, 1956, p. 42) was used. Such procedures were necessary as a random sampling of teachers and schools was deemed impractical in view of the nature of the study.

Analyses

The facilities of the Cornell Computing Center were employed for most of the analyses (Control Data 1604). A special program, developed by L. Wightman (1965), a Cornell University graduate student, was modified by him to fit the needs of the project. The program plotted matrices, calculated I-D ratios and percentages, grouped matrices, performed Darwin tests, and did Scott reliability checks. Correlational and regression analyses were carried out with standard Cornell "Custat" programs (Searle, 1964a, 1964b, 1964c).

In addition to analyzing teachers on the basis of total instruction, matrices were produced for each teacher's interaction to determine influence for each of four "major activities." These were: (1) Planned Demonstration, (2) Lecture, (3) Laboratory, and (4) Recitation. A grand matrix (composite) of all teachers for each Major Activity and for Total Instruction was constructed. Other composite matrices of interest were formed as well (e.g., a composite for each teacher's total teaching; fall totals and spring totals for each teacher based on Total Teaching and for each Major Activity).

To investigate "directness" versus "effectiveness" and "flexibility" versus "effectiveness," teachers were grouped as "high" and "low" on the independent criteria. For each aspect of effectiveness, effectiveness scores (adjusted class achievement scores) of the upper third teachers ("high," N = 5) were compared to those of the lower third teachers ("low," N = 5) using the Mann-Whitney U Test. The null hypothesis employed was "H_0 : There is no difference between the effectiveness of teachers 'high' on the independent criterion and the effectiveness of teachers 'low' on the independent criterion." The corresponding alternative hypothesis was "H_1 : The effectiveness of 'high' teachers is stochastically greater than that of teachers 'low' on the independent criterion."

In adjusting test scores for the above comparisons, linear multiple regression analyses were conducted to determine the best combination of variables yielding a maximum number of students having complete sets of scores on all variables ultimately used.

Numerous Pearson product-moment and Kendall tau correlations were carried out to discover correlating variables and to identify correlates connecting effectiveness with aspects of the Flanders matrix and with various teacher and school characteristics.

Concluding Remarks

It should be mentioned that extensive efforts were made to (1) develop techniques for employing the system in such a way that the entire teaching episode is "captured," (2) determine reliability of the observation data, (3) assess teacher and student reaction to the observational situation, (4) maintain superior rapport with participating teachers and schools, and (5) to provide participating teachers and schools with information resulting from the project that may be of some use to them.

CHAPTER FOUR

THE SAMPLE OF PARTICIPATING TEACHERS AND SCHOOLS

Introduction

In designing the study, a pragmatic decision was made at the outset to limit observations to a representative sample of physics teachers teaching the New York State Regents Physics Course in public high schools. The limitation to public high schools provided additional control for the effectiveness phase of the study. The assumption that physics teaching in these classrooms is typical of high school physics teaching in general is reasonable.

In view of the nature of the study, it was felt that a random sampling of teachers and schools would not be feasible. Consequently, the sample was constructed through solicitation in such a way that it adequately represented a defined population of teachers and schools according to selected characteristics of both.

Sample Size

The experimental design called for three daily fall observations (consecutive days), and similarly, three for the spring. The work of Flanders has shown that after 180 minutes of observation data on a teacher, the distribution of tallies in matrix cells becomes essentially stable. For this reason, it was felt that at least six data-productive teaching periods on each teacher would be necessary. By doing half the observations in the fall and half in the spring, a teacher would be observed undertaking each of two different areas of subject matter. Also, comparisons of fall and spring composite matrices would provide an indication of how well the matrix data could be associated with consistent behavioral aspects of the teacher's performance (teaching style).

One objective was to produce a grand composite matrix of secondary physics teaching which could be compared to composites reported by Flanders. With at least fifteen teachers carried through the study, a matrix on the order of 80,000 tallies (about 100 periods of teaching) would result. This would be more than enough to provide a good comparison to the reported composites for seventh grade social studies and eighth grade mathematics (each less than 60,000 tallies).

Other objectives concern teacher effectiveness. It was felt that with fifteen teachers' matrix scores available, the ranges of the latter would be reasonably extensive; i.e., it seemed likely that there would be, for example, some very direct and very indirect teachers within the sample. This would, of course, enhance the planned comparisons (we cannot see which "kind" of teacher is more effective if the sample members are, in the main, all of one "kind").

A margin of two teachers was required for the event of a teacher's dropping out of the study. With a total of seventeen, then, the fall and spring series of observations could be conducted in ten weeks each, providing some of the schools were close enough to each other so that most of the time, two schools per week could be observed. (An additional controlling condition was to observe each teacher on Monday through Thursday in the fall, and Tuesday through Thursday in the spring.)

Finally, seventeen teachers in seventeen schools appeared to be the upper limit from the practical standpoint. It was doubted that more than seventeen would enhance the study enough to make the extra work and expense worthwhile. It was thus decided to construct a sample of seventeen teachers representing seventeen high schools.

The Sampling Technique Used

Reasons for Non-Random Sampling

The nature of the research required that the following conditions be met to enhance the quality of the data:

1. It was important that participating teachers were unquestionably willing to take part in the study; that they did not feel uncomfortable about the observer's presence in the classroom (fear of being judged, jeopardy of their positions, etc.); that they had full confidence in the observer's professional

ethics; and that they just taught "the way they ordinarily do."

2. It was necessary that administrators and supervisors be happy with the arrangement and willing to make students' permanent records available to the investigator.
3. Teachers had to agree to: (a) administer two tests, both pre and post, therefore "lose" four class periods; (b) not give a test longer than ten minutes, nor show a long movie, nor take a field trip during fall and spring observations. Otherwise, they were to do whatever they ordinarily did.
4. Schools had to have a pairwise geographical proximity (twenty or thirty miles, say) making it possible to observe two schools in one day in most cases. (Pairs were scattered.)

It was clear that with conditions such as these a random sampling of schools was not practical.

The Population Defined

As a random sampling of schools and teachers was not practical, it appeared better to confine a population to a chosen region of New York State and to limit predictions to such a group of schools. Accordingly, the population was defined as the group of all public high schools located within the union of the following fourteen counties of New

York State surrounding Ithaca, New York: Broome, Cayuga, Chemung, Chenango, Cortland, Madison, Onondaga, Ontario, Schuyler, Seneca, Steuben, Tioga, Tompkins, and Yates. This union contained 131 public high schools and is referred to as a "central New York State region."

Selected Population and Sample Characteristics

A sample of seventeen schools (one teacher in each school) was constructed in such a way that it adequately reflected the population on twelve pre-selected characteristics. Characteristics were selected on somewhat of an *a priori* basis, i.e., the ones chosen just seemed to provide a basis upon which to construct a sample of schools. The following outline defines the characteristics employed.

A. High School Size As Indicated By Enrollment

A_1 : Grade 10 enrollment.

A_2 : Grade 11 enrollment.

A_3 : Grades 10..12 enrollment.

B. The Physics Class

B_1 : Regents physics class size. (If more than one class in a school, the average class size was used.)

B_2 : Periods per week of Regents physics instruction. (Includes laboratory.)

B_3 : Percent physics enrollment is of grades 10..12 enrollment. (Other physics courses used in place of the Regents course were included.)

C. The Physics Teacher

C_1 : Years of experience in secondary science teaching.

C_2 : Total years of teaching experience in any subject at any level.

C_3 : Credit hours of collegiate physics preparation.

C_4 : Credit hours of collegiate science preparation. (Reported as the sum of biology, chemistry, physics, and earth science.)

D. The School District

D_1 : Expenditures per pupil in average daily attendance in the district for 1961-1962. (The district figure is cited for each high school in cases of districts having more than one high school; given in dollars.)

D_2 : Ratio of district's real property full valuation for tax purposes and levy, 1961-1962, to district's total K..12 enrollment. (The same ratio is cited for each high school in cases of districts having more than one high school; in thousands of dollars per student.)

The above selection is, of course, not entirely *a priori*. For example, school size might be associated with facilities, with urban areas,et cetera. Schools "high" on B_3 might place more emphasis on a secondary science sequence than those who are "low"--or interest in college entrance might be more extensive. Teacher experience and subject matter background are associated with preparation for teaching. Section D characteristics likely relate to community socio-economic level. These are, at best, speculations.

The rationale used was that if, for each of these characteristics, the distribution of values in the sample was "sufficiently similar" to that of the population, then the sample would be statistically as good as a random sample and generalization of findings to the population would be possible. The notion "sufficiently similar" will be clarified shortly.

Data for the selected characteristics in groups A, B, and C were obtained from the "Brown Card" file of the Office of Science Education of the New York State Department of Education, Albany, in August 1964.[1] These data were reported in the fall of 1963. Once sample schools were obtained, fall 1964 data were gotten directly from the participants. The source of data on D_1 and D_2 was the Annual Educational Summary, New York State, 1961-1962 by the Bureau of Statistical Services of the State Education Department.

For each of the twelve selected characteristics, a frequency distribution was plotted (Appendix A, Figures A-1 through A-12). In each case a plot of 1964-1965 sample data is superimposed upon a plot of 1963-1964 population data.

Table 1 summarizes the population data. It is to be noted that the standard deviation used here (and throughout the study) is the positive square root of the unbiased variance estimate:

$$S_X = \sqrt{\frac{\sum_i (X_1 - \bar{X})^2}{N - 1}} = \sqrt{\frac{N}{(N - 1)}\left(\overline{(X_1^2)} - (\bar{X})^2\right)}$$

---

[1] The author is grateful for the assistance of Mr. Robert MacGregor, Associate in Science Education, State Education Department.

TABLE 1

STATISTICS ON SELECTED CHARACTERISTICS OF 131 PUBLIC HIGH SCHOOLS COMPOSING A 14-COUNTY CENTRAL NEW YORK STATE REGION

Characteristic[a]	No. Schls. Providing Data (N)	N/131	$\bar{X}$	Mdn.	S_X	$S_{\bar{X}}$	$\bar{X} \pm 2S_{\bar{X}}$
A_1	111	0.85	146	90	135	12.8	146 ± 26
A_2	110	0.84	137	80	133	12.4	133 ± 25
A_3	131	1.00	414	275	389	34	414 ± 68
B_1	111	0.85	16.8	18	7	0.7	16.8 ± 1.4
B_2	107	0.82	6.3	6	0.8	0.1	6.3 ± .2
B_3	100	0.76	7.9	7.2	4.1	0.4	7.9 ± .8
C_1	122	0.93	8.5	6	8.5	0.8	8.5 ± 1.6
C_2	121	0.92	9.3	6	9.2	0.8	9.3 ± 1.6
C_3	122	0.93	23.2	18	14.5	1.3	23.2 ± 2.6
C_4	122	0.93	71.3	71	22.2	2.0	71.3 ± 4.0
D_1	131	1.00	536	535	52	5	536 ± 10
D_2	131	1.00	15.8	13	8.1	0.7	15.8 ± 1.4

[a] For definitions of characteristic designations, see pages 47 and 48.

The standard deviation of the sampling distribution of means is estimated by:

$$S_{\bar{X}} = S_X / \sqrt{N}$$

These data were considered carefully in reaching a decision concerning which schools to solicit for participation. It was decided that each of the fourteen counties would be represented and that several cities in addition to non-urban schools would be included. Consequently, participating schools were scattered about the region. One female teacher was sought as 1-to-17 is typically the female-to-male physics teacher ratio. Similarly, one class of the new Experimental Regents Physics Course (Van Hooft, 1964) was required (especially in connection with Objective One.)

School and Teacher Solicitation Procedure

After identifying "potentially desirable" participants, an appropriate letter of solicitation was sent to school administrators. A copy of it was included with a letter to teachers in another mailing (Appendix B, Forms B-1 and B-2, respectively). A brief indication of the nature of the study was included with a request for a personal meeting with the principal and teacher (together). Appointments were set up by telephone and the meetings were held at the schools. It was at these meetings that solicitations for participation in the study were carried out.

In each case, copies of a "Memorandum to Physics Teachers, Science Department Heads, and Principals" was distributed (Appendix B, Form B-3). The latter outlined details of what the investigator required of teachers and schools, as well as what services he would render in return, should they decide to participate. Only nineteen solicitations were required to obtain a sample of seventeen. The willingness of teachers and administrators to cooperate with the study was very encouraging.

Comparing the Sample to the Population

As anticipated, no school reflected the population as a whole. However, the sample reflected the population very well on the basis of the twelve pre-selected characteristics described earlier. For each of the latter, 1964 sample data were compared to 1963 population data using the Kolmogorov-Smirnov One-Sample Test. It was assumed that the change in mean values (and distribution) in the population data, in one year, would be negligible.

The Kolmogorov-Smirnov Test (non-parametric) provides a measure of goodness of fit, i.e., it indicates the degree of agreement between the distribution of a set of sample values and a specified (theoretical) distribution. In particular, it "determines whether the scores in the sample can reasonably be thought to have come from a population having the theoretical distribution" (Siegel, 1956, p. 47). In this case the theoretical distribution (population) was known. The statistic is a calculated maximum divergence between the two distributions. Its sampling distribution indicates the probability of a divergence as large as the one observed if the sample were really a random sample from the population. For each characteristic tested, the null hypothesis was:

> H_0 : There is no difference between the sample and population distributions of the criterion.
> alpha = 0.10

It is noted that:

$$D_{max} = \text{maximum} \left| F_0(X) - S_N(X) \right|$$

where $F_0(X)$ is the theoretical cumulative distribution (population) under H_0 and $S_N(X)$ is the observed cumulative frequency distribution. In these tests, $N = 17$. In effect, the tests were done to see if S_{17} was that of a random sample. Under H_0, values of D_{max} of 0.250 and 0.286 would have probabilities of occurrence of .20 and .10 respectively.

Table 2 summarizes the results. It was not possible to reject H_0 at the .10 level on any of the twelve characteristics. On most criteria, the probability of D_{max} as large as the observed value was on the order of .5 or higher. It was thus concluded that the sample was "sufficiently similar" to the population to facilitate generalization of findings to the latter, i.e., that the sample was a random sample of the population.

TABLE 2

COMPARISONS OF SAMPLE AND POPULATION DISTRIBUTIONS OF 12 PRE-SELECTED CHARACTERISTICS OF TEACHERS AND SCHOOLS USING THE KOLMOGOROV-SMIRNOV ONE-SAMPLE TEST[a]

Characteristic[b]	Number in Population	Number of Intervals	D_{max}[c]	Probability of D_{max} Under H_0
A_1	111	29	.174	$>.20$
A_2	110	32	.265	about .15
A_3	131	36	.227	$>.20$
B_1	111	19	.173	$>.20$
B_2	107	5	.128	$>>.20$
B_3	100	27	.241	about .25
C_1	122	38	.207	$>.25$
C_2	121	43	.171	$>.20$
C_3	122	35	.186	$>.20$
C_4	122	34	.134	$>>.20$
D_1	131	34	.087	$>>.20$
D_2	131	28	.147	$>>.20$

[a] Sample, N = 17; population, N = 131.

[b] For definitions of characteristics designations, see pages 47 and 48.

[c] Critical values of D_{max} are .250 and .286 at the .20 and .10 levels respectively (one-tailed test).

An alternative approach was carried out as a matter of interest, and perhaps, to provide additional support to the conclusion. Table 3 shows that for all criteria the population mean was within the 95% confidence limits of the value predicted by the sample. Only 6 of 17 criteria met this condition if 68% confidence limits were substituted.

Finally, Table 4 shows values of other characteristics of the sample. These are given to provide a picture of the sample which is a bit more complete.

Concluding Remarks

An extensive effort was made to show that the sample of participating teachers and schools adequately represented a defined population. This was necessary to allow generalization of findings.

It is interesting to note that selected characteristics in the population of 131 schools were not normally distributed (Appendix A). Parametric tests of significance generally assume normality of criteria of interest in the population. Non-parametric tests do not. Although a little less powerful than their parametric counterparts, non-parametric measures are often more appropriate as illustrated in the problem of this chapter.

In January 1965 a "Summary of Data of School Characteristics" report was submitted to participating teachers and administrators as a service (see Appendix B, Forms B-4 and B-5).

TABLE 3

STATISTICS ON SELECTED CHARACTERISTICS OF THE SAMPLE OF 17 SCHOOLS COMPARED TO THE POPULATION OF 131 SCHOOLS[a]

Characteristic Designation	$\bar{X}$	Mdn.	S_X	$S_{\bar{X}}$	95% Confidence Limits on Mean $\bar{X} \pm 2S_{\bar{X}}$	Population $\bar{X}$ within Prediction? 68%	95%
A_1	184 (146)	119 (90)	142 (135)	34.4 (12.8)	184 ± 69 (146 ± 26)	no	yes
A_2	174 (137)	115 (80)	137 (133)	33.2 (12.4)	174 ± 64 (133 ± 25)	no	yes
A_3	523 (414)	338 (275)	413 (389)	100.2 (34.0)	523 ± 200 (414 ± 68)	no	yes
B_1	18.4 (16.8)	19 (18)	6.24 (7.00)	1.51 (0.70)	18.4 ± 3.0 (16.8 ± 1.4)	no	yes
B_2	6.4 (6.3)	6 (6)	0.7 (0.8)	0.2 (0.1)	6.38 ± 1.41 (6.3 ± 0.2)	yes	yes
B_3	8.2 (7.9)	8.1 (7.2)	1.6 (4.1)	0.4 (0.4)	8.18 ± 0.76 (7.9 ± 0.8)	yes	yes
C_1	6.4 (8.5)	4 (6)	7.9 (8.5)	1.9 (0.8)	6.35 ± 3.82 (8.5 ± 1.6)	no	yes
C_2	7.0 (9.3)	5 (6)	7.1 (9.2)	1.7 (0.8)	7.00 ± 3.46 (9.3 ± 1.6)	no	yes
C_3	25.2 (23.2)	24 (18)	12.7 (14.5)	3.2 (1.3)	25.2 ± 6.36 (23.2 ± 2.6)	yes	yes
C_4	76.1 (71.3)	79 (70)	21.1 (22.2)	5.1 (2.0)	76.1 ± 10.24 (71.3 ± 4.0)	yes	yes
D_1	538 (536)	527 (535)	65 (52)	16 (5)	538 ± 31.4 (536 ± 10)	yes	yes
D_2	18.5 (15.8)	13.7 (13)	11.8 (8.1)	2.9 (0.7)	18.5 ± 5.7 (15.8 ± 1.4)	yes	yes

[a] In each case, upper number corresponds to 1964-65 sample data and lower number corresponds to 1963-64 population data.

TABLE 4

ADDITIONAL CHARACTERISTICS OF THE SAMPLE OF 17 PARTICIPATING HIGH SCHOOLS, 1964-1965

Characteristic	$\bar{X}$	S_X	$S_{\bar{X}}$	95% Confidence Limits
Teacher age, 9/64, to nearest year	32.1	7.75	1.88	32.1 ± 3.8
Number of Daily Preparations	2.5	1.19	0.29	2.5 ± 0.6
Teacher Load, Teaching Contact Hours Per Week	20.1	2.49	0.60	20.1 ± 1.2
Teacher Load, Non-teaching Contact Hours Per Week	2.4	1.72	0.42	2.4 ± 0.8
Teacher Load, Total Contact Hours Per Week	22.5	2.76	0.67	22.5 ± 1.3

CHAPTER FIVE

SYSTEMATIC OBSERVATION IN THE HIGH SCHOOL PHYSICS CLASSROOM - COLLECTION OF DATA

The Observation Instrument

Flanders' Categories

Flanders' categories for interaction analysis were employed. Three basic points merit emphasis. (1) The major purpose of the Flanders' categories is the analysis of the teacher's verbal behavior in the classroom. (2) The behavioral dimension upon which the categories are based is directness (or indirectness) of teacher influence. Research has shown this to be a "most promising" dimension. (3) The observational scheme is systematic. It is an objective method of acquiring data on teacher behavior. Observer judgment is involved only to the extent that the observer decides into which of ten categories the dominant type of behavior of a three-second interval belongs.

The ten mutually exclusive categories for coding teacher-student communications in the classroom were considered to be essentially inclusive of all verbal teacher-student discourse. Communication was categorized in three-second intervals with the condition that more than one categorization within such an interval would be recorded should a shift in category designation take place.

In the following outline, Flanders' category definitions are quoted (Flanders, 1964a, p. 5).

CATEGORIES FOR VERBAL INTERACTION ANALYSIS[1]

TEACHER TALK	INDIRECT INFLUENCE	1.[2] **ACCEPTS FEELING:** accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.
		2. PRAISES OR ENCOURAGES: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying "um hm?" or "go on" are included.
		3. ACCEPTS OR USES IDEAS OF STUDENT: clarifying, building, or developing ideas suggested by a student. As a teacher brings more of his own ideas into play, shift to category five.
		4. ASKS QUESTIONS: asking a question about content or procedure with the intent that a student answer.
TEACHER TALK	DIRECT INFLUENCE	5. LECTURING: giving facts or opinions about content or procedure; expressing his own ideas, asking rhetorical questions.
		6. GIVING DIRECTIONS: directions, commands, or orders to which a student is expected to comply.
		7. CRITICIZING OR JUSTIFYING AUTHORITY: statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.
STUDENT	TALK	8. STUDENT TALK--RESPONSE: a student makes a predictable response to teacher. Teacher initiates the contact or solicit student statement and sets limits to what the student says.
		9. STUDENT TALK--INITIATION: talk by students which they initiate. Unpredictable statements in response to teacher. Shift from 8 to 9 as student introduces own ideas.
		10. SILENCE OR CONFUSION: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

---

[1] From: Ned A. Flanders, Interaction Analysis in the Classroom, A Manual for Observers (Ann Arbor, Michigan: University of Michigan, School of Education, January 1964).

[2] There is no scale implied by these numbers. Each number is classificatory, it designates a particular kind of communication event. To write these numbers down during observation is to enumerate, not to judge a position on a scale.

The following illustrative remarks and examples are added to clarify the definitions, as used in this study, in terms of the high school physics classroom setting. It is noted that Categories 1 through 7 account for teacher talk (1 through 4 are indirect influence; and 5 through 7, direct influence). Categories 8 and 9 deal with student statements and Category 10 is "Silence or Confusion." The definitions are repeated for emphasis.

> CATEGORY 1. ACCEPTS FEELING: accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings are included.

Remarks: In the high school physics classroom, practically all of student feeling is associated with ease or difficulty, and failure or success, in grasping subject matter concepts and in problem solving. Occasionally, feelings associated with classroom procedures are involved. Teacher recognition of feelings leads to "1's."

Examples of Teacher Statements in Category 1:

"I think I see what's confusing you in this problem."

"I know you have trouble with these, but. . . ."

"I think you will find this interesting."

"I remember that you were unhappy with last week's assignment."

"Apparently the notion of a proportionality constant is confusing to you."

CATEGORY 2. PRAISES OR ENCOURAGES: praises or encourages student action or behavior. Jokes that release tension, not at the expense of another individual, nodding head or saying "um hm?" or "go on" are included.

Remarks: In the high school physics classroom, there is very little teacher praise. Practically all praise and encouragement is in reference to student behavior as a "physicist," i.e., in working with secondary physics subject matter. Teachers occasionally respond with a praise statement when a student gives the right answer, solves the problem correctly, makes a good suggestion, etc. Extended praise is essentially non-existent. Teacher-provided subject matter clues to encourage student response are included.

Examples of Teacher Statements in Category 2:

"Very well done, John."

"That is a fine suggestion."

After asking a student to calculate the current through one of two resistors in parallel, and getting no response, the teacher says, "Now the voltage will be the same across both resistors, won't it?"

"Come on, I know you can do it."

CATEGORY 3. ACCEPTS OR USES IDEAS OF STUDENT: clarifying, building, or developing ideas suggested by a student. As a teacher brings more of his own ideas into play, shift to category five.

Remarks: Clarification is to be distinguished from the verbal habit of repeating what the student says. Most teachers repeat nearly every student answer. If there is evidence, from the teacher's voice tone and the content of the student's statement, that the teacher has accepted the student's statement or idea, a "3" applies. This holds whether the teacher's response is a repeat of the student's statement or not. Getting a student to repeat part of his statement(s) is also a "3." In the high school physics classroom extended building of student ideas, beyond a minute, say, is uncommon. As soon as the teacher terminates direct usage of the student's idea or statement, usually done by a content or opinion statement or by posing another question, other appropriate categories apply.

Examples of Teacher Statements in Category 3:

Student gives an answer, "Two watts"; and the teacher says, "Two watts. Right." or just "Okay." (accepting)

"Did you mean that the current through this resistor is one amp?" (clarifying)

"You're saying the velocity change is 3 feet per second in a time of 0.5 seconds, so you mean then that the acceleration is 6 feet per second squared." (clarifying)

Student says, "Two grams times three feet per second"; and the teacher, unsure of what the student said, responds with "how many grams did you say?" (clarifying)

"Let us see what we get if we use Bill's value of 2.3." (building)

Student suggests a proportionality between stretching force on a spring and elongation. Teacher says, "Let's plot force versus elongation and see what we get"; or teacher says, "If they are proportional, we can write an equation using a constant of proportionality." (building)

CATEGORY 4. ASKS QUESTIONS: asking a question about content or procedure with the intent that a student answer.

Remarks: Interrogative statements of the teacher could conceivably fit any of the seven teacher talk categories, but any teacher solicitation of information from the student concerning content or procedure would generally be classified as a "4." Declarative teacher statements, similarly soliciting, would also fit this category. Rhetorical questions are not included; a teacher question immediately followed by a statement belonging to another teacher category is not considered as a question for there is apparently no intent that a student answer. A question on content or procedure followed by at least a three-second pause is categorized as "4" whether there is a student response or not. A short teacher statement of fact or opinion (one or two sentences) followed by the inquiry "right?" or "true?" in which the teacher solicits a student response is considered as a question. The observer must be alert to the verbal habit of a "right?" after declarative statements in which student response is not solicited. Such instances do not fit Category 4.

Examples of Teacher Statements in Category 4:

"What is Newton's First Law?"

"How did you set up your apparatus?"

"Give me your answer to problem six."

"The current through this resistor will be three amps. Right?" (Teacher pauses for student acknowledgement.)

"What do you think the relation between force and elongation might be?"

"Where do we go from here?"

"Did you bring your slide rules?"

CATEGORY 5. LECTURING: giving facts or opinions about content or procedure; expressing his own ideas, asking rhetorical questions.

Remarks: In the high school physics classroom, this is the most commonly used category. The teacher's presentation of facts, opinions, or his own ideas accounts for most of the tallies. A teacher invitation for a student to repeat his entire statement (the teacher did not hear what the student said), such as "pardon me," or "what?" is included. When a teacher reads aloud, this category applies unless directions are read.

Examples of Teacher Statements in Category 5:

"Ohm's Law is I = E/R."

"The answer to problem six is 17 feet."

"Let's look at problem six again."

"Tomorrow we will have a test on electricity."

"Force is directly proportional to displacement. What units are associated with the constant of proportionality? Well, we could have dynes per centimeter or pounds per foot."

"In your text on page ten, question two reads, 'What are Kirchoff's Laws?' These laws will be useful to us in solving DC circuit problems."

"The energy of a light quantum is proportional to frequency."

CATEGORY 6. GIVING DIRECTIONS: directions, commands, or orders to which a student is expected to comply.

Remarks: Directions related to classroom management and procedures are conscepicous as "6's." However, an invitation which is not a command would not fit this category. For example, "Do problem ten," is clearly a "6" but, "You might like to do problem ten, " would be a "5." The latter could possibly be a "2" (encouragement). Often a decision is made in terms of the setting of the situation.

Examples of Teacher Statements in Category 6:

"Take your seat."

"Turn to page sixteen."

"Set up a series circuit using the 6-volt battery, the 1200 ohm resistor, and the milliammeter."

"Check your calculations again."

"Use 10, 20, and 30-gram weights."

CATEGORY 7. CRITICIZING OR JUSTIFYING AUTHORITY: statements intended to change student behavior from non-acceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.

Remarks: The essence of the criterion for this category is teacher criticism of student social behavior and of student behavior as a "physicist," i.e., student behavior in the handling, use and consideration of subject matter content and ideas. In high school physics classes practically all of teacher statements in this category are associated with the latter situation. Although both teacher directions and criticism will be connected with a change in student behavior (unacceptable to acceptable) only criticism fits Category 7. Self-reference statements are included if there is evidence that the teacher is dissatisfied with current student behavior. Vocal intonation as well as statement content serve as important observer cues. Whenever a teacher rejects part or all of a student communication, i.e., says in effect, "No, that is wrong," a "7" is clearly appropriate.

Examples of Teacher Statements in Category 7:

Student gives an answer and teacher says, "Un hun," or "No."

"You are not handling the equation properly."

"Why didn't you do your homework?" (harshly)

"No, force is directly proportional to mass."

"I'm the one to decide this and what I say goes!"

CATEGORY 8. STUDENT TALK--RESPONSE: a student makes a predictable response to teacher. Teacher initiates the contact or solicits student statement and sets limits to what the student says.

Remarks: Only student-to-teacher communication fits this category. Teacher-imposed limitations, predictability of student response and demonstration of student initiative are key factors in classifying student statements as "8's." The observer must categorize on the basis of whether the teacher is expanding or restricting freedom of action. Restriction corresponds to "8." Student responses to "narrow" questions and brief answers to problems (usually numerical--what the student "got") apply whether the student indicated a desire to respond (hand raising) or not. A wrong answer to a predictable response question (narrow) is also included. In cases of student explanation of an answer or solution, student initiative would guide the observer. Little or no initiative would be cause for "8."

Examples of Student Statements in Category 8:

"Force equals mass times acceleration."

"The answer to problem eight is 23 grams."

"I got the same answer the book gives."

"I hooked the resistors up in series like it says in the lab manual."

Teacher criticizes student for failure to pay attention. Student says, "I'm sorry," or "Okay."

"I used the formula f = ma and put in the values given after changing kilograms to grams." (The decision on student initiative, in this case, would likely be affected by a "restrictive" or "expanding" interpretation of the setting at the moment. The nature of the question producing the response would be considered.)

CATEGORY 9. STUDENT TALK--INITIATION: talk by students which they initiate. Unpredictable statements in response to teacher. Shift from 8 to 9 as student introduces own ideas.

Remarks: Both student-to-teacher and student-to-student communication fits this category. The key factors of

Category 8 apply, i.e., if little or none of teacher limitation and of response predictability, but definite involvement of student initiative, a "9" would be appropriate. All student questions and student-to-student statements where the teacher is a participant (active or passive) are "9's." If the teacher is not participating in any way, student talk is ignored. Answers to "broad" questions fit here, although early portions of the response may fit Category 8. When a student reads material not his own, Category 8 applies. When the material is his own, "9's" would be appropriate if the student was presenting his own ideas, showing initiative, et cetera. "Borderline" cases are often settled in terms of whether the activity at the moment is teacher "restrictive" or "expanding."

Examples of Student Statements in Category 9:

"What is the answer to problem six?" (to teacher)

Teacher solicits suggestions concerning a problem. Student offers, "Why not consider f = ma?"

"I think Bill has the wrong idea."

"Let's use just one spring but let's try hanging different masses on it and see how the stretch varies with mass."

Teacher criticizes student for inattentiveness; student retorts, "I wasn't doing anything."

CATEGORY 10. SILENCE OR CONFUSION: pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

Remarks: Student-to-student communication in which the teacher is neither an active nor passive participant is ignored. If no other communication is taking place, "10" would apply. The observer always directs his attention to the teacher. Any three-second pause in the communication is a "10." After thirty seconds of "10's," the observer stops recording until categorizable communication resumes.

It should be emphasized that although categorizations were done at a rate of twenty per minute (every three seconds), a category shift within an interval would be considered. For example, a rapid teacher-student-teacher exchange consuming three seconds in its entirety, would yield three tallies for that interval. Also, in the event of student-to-student communication (teacher participating), "10's" would be inserted between sequences of tallies of individual student contributions.

With only the defined categorization scheme considered, numerous "borderline" occurrences of questionable coding arise. Flanders' General Ground Rules were followed; however a special set of "ground rules" for the physics teaching in this study had to be developed. These efforts were necessary to enhance observer reliability.

General Ground Rules Used

Several general ground rules (see Amidon and Flanders, 1963; and Flanders, 1964a, for a complete description of the basic system) were used. Two of the rules are the essence of the preceding paragraph. It is best to quote the first three below:

> Rule 1: When not certain in which of two or more categories a statement belongs, choose the category that is numerically farthest from Category 5.
>
> Rule 2: If the primary tone of the teacher's behavior has been consistently direct or consistently indirect, do not shift into the opposite classification unless a clear indication of shift is given by the teacher.
>
> Rule 3: The observer must not be overly concerned with his own biases or with the teacher's intent. . . . The effect of a statement on the pupils, then, not the teacher's intent, is the crucial criterion for categorizing a statement. (Amidon and Flanders, 1963, pp. 21-23)

Special Ground Rules Used

For the purposes of this study, it was necessary to develop a set of "special ground rules." The detailed report given here and in the preceding sections is necessary to assist proper comparisons of the results of this study and others using the Flanders method (*viz*., Matthews, 1965) and to enhance any possible replication efforts in the future. Also, the "special ground rules" serve to further clarify applied procedures and interpretations of the category definitions.

In the outline it is noted that "0" means "10." By writing one digit for all categories, it is easier for the observer to maintain accurate timing. Also, "T" denotes "teacher" and "S" means "student."

Outline of Special Ground Rules[3]

1. Additional Communication Not Coded

   a. S-to-S discourse with T not participating in any way.

   b. Routines. Any announcements pertaining to activities not associated directly with the discourse of the lesson or major class activity is "routine." For example, T announcing a National Science Foundation program that S's may apply for, announcing an assembly program, public address system announcements, daily announcements from the office.

   c. T announcement of assignment at beginning of period is not coded. But any discussion of assignment and/or announcement within Major Activity (Demonstration, Lecture, Recitation, Laboratory) is coded.

   d. T dictation of questions for a class quiz.

2. When Teacher Poses Problem

   a. When posing problem (for lab, say), even though a declarative statement is used, code is 4. For example, T: "Determine the index of refraction." But specific directions, viz., T: "Stick the pin in the board so that it makes an incident angle with the face of the prism (not 0 or 90 degrees) and then look through the prism and locate the pin" yields 6's.

   b. If T poses problem (viz., T reads problem) and then pauses to allow S response, then posing is a 4. But if T proceeds to answer without pause for S comment, then a 5 applies.

   c. If T presents "open" question which S's are to "think about," then the category is 4.

3. Teacher Terminates With "Any Questions?"

   a. If then T pauses and allows time for response,

---

[3] In this outline the designation x,y,z implies transitions from category x to category y to z. There can be any amount of consecutive time intervals in x, y, or z (e.g., x,y,z is 4,4,4,8,8,3,3,3,3,3,3,3).

"any questions," or its equivalent, this is a 4. Example: "Any questions on this? (pause) All right, let's go on." This would be logged as 4, 0,5. But, "Any questions on this? Now acceleration is time rate of change of velocity" would be 5,5.

4. Teacher Provides Clues to Student

a. T asks question. S does not respond immediately, but hesitates. T gives additional information as a clue (as seen by S). Such a clue is considered ENCOURAGEMENT. Sequence would be 4,0,2. But if T just repeats question or asks someone else, code is 4,0,4.

b. If S responds to question with something correct, but not what T was looking for, may have 4,8-9, 3,4. Example: T says, "Why does it increase?" S says something acceptable, but not what T wanted. T says, "Yes, . . . but why does it increase?"

   1) But, if in this case, T says "Yes" and then instead of repeating the question, gives information which would clue S to the answer sought, the sequence would be coded 4,8-9,3,2.

   2) If S's remark is unacceptable, and T rejects it and then repeats the question, have 4,8-9,7,4.

   3) If T rejects and then clues, would be 4,8-9,7,2.

   4) S response is unacceptable, but T ignores it and repeats question, code is 4,8-9,4.

c. With a severe shift in category, rejection, 7, and clueing,2, would have to be very obvious. If questionable as a clue, the T's remark might go down as a 5.

5. Student Response Not An Answer

a. T asks a content question. Any S response not the right answer nor wrong answer would be a 9. Example: S says, "I don't know the answer" or "Huh?" or "What did you say?"

6. Teacher Verbal Habit of "Right?"

   a. If done after T statement, a "Right?" is a 5 if no pause follows. ("Okay?" is equivalent to "Right?")

   b. If definite pause but no response from S's, and then T goes on, sequence is 4,0,5; i.e., "Right?" is a 4.

   c. T says, "Right?" and S(s) respond, "Yes," the code is 4,8. (A nod is "Yes.")

   d. T says "Right?" and S(s) respond with question or something other than "Yes," code is 4,9.

7. When Teacher Reads Aloud

   a. T reads directions, code is 6.

   b. T reads question to be answered, code is 4.

   c. T reads problem and proceeds to solve it himself; then code is 5.

8. Category 7 Versus Category 3

   a. When T refers to what S says (e.g., "John says . . . ."), 3 applies providing T is accepting or clarifying. But the code could be 7 if T is obviously critical of S's remarks.

   b. T question about what S said can be criticism, 7. Example: T says, "Newton's first law is F equals ma?" If tone is of disbelief, or condemnation, S hears in effect, "You are wrong!" This is a 7. But if the inquiry is to the effect, "Is this what you are saying?" then T is clarifying and 3 applies. It is noted that T can clarify without accepting.

   c. The type of T query in "a." or "b." above is not a 4. If not obviously 3 or 7, it is left as 5. This is not an exception to General Ground Rule 1 because 3 and 7 are equidistant from 5. It may depart from General Ground Rule 2.

9. Breaks in Communication

   a. When class is interrupted by an "outside factor" (*viz.*, another T comes in to borrow something, or the public address system goes on) the observer puts down two O's and stops recording. This holds regardless of how long the interruption continues.

b. If disruption within the class, viz., S misbehavior, and no T verbalizations, O's recorded to a maximum of 10 (30 seconds). A S-T or T-S exchange would end the "break."

c. If pause at end of activity, but termination not obvious, maximum of 10 O's recorded.

d. In lab situation, as T goes from S to S, group to group, or from table to table, only two O's are recorded between each sequence of exchanges.

Some of the items in the above outline may appear "unnecessary" in light of stated category definitions. The outline is given as it was used--as a guide. It was developed as troublesome instances arose.

Observer Training Program

All systematic observations were carried out by the Principal Investigator. The observer received his initial training in the Flanders method of interaction analysis in the course of a two-day workshop at Cornell University in early spring, 1964. The workshop was conducted by Dr. Edmund Amidon of Temple University and was sponsored by Cornell's Junior High School Project. About twelve hours of instruction were involved.

Considerable practice in coding from tape recordings was acquired in ensuing days. Master tapes from Amidon were used; transcripts of taped discourse were written out and analyzed.

"Live" coding in several classrooms was done on a "pilot" basis. In this connection, classes visited included second grade, ninth grade English and social studies, high school chemistry and college physics. Tape recordings were made and the "live" data were compared to that obtained from the tapes.

Numerous conferences were held with two other observers also interested in the techniques. Interpretations of category definitions were discussed at length and reliability measures taken. Techniques for recording and coding were developed.

Observation Techniques

Schedule of Classroom Visits

Each of the seventeen participating classes (seventeen high schools) was observed for four consecutive days (Monday through Thursday) in the fall "series of observations" and for three consecutive days (Tuesday through Thursday) in the spring. Fall observations began in mid-October 1964 and were completed in mid-December. Spring observations ran from mid-February through mid-May 1965. With few exceptions, the same order of visits conducted in the fall was used in the spring. Exceptions involved several schedule changes due to such things as an unexpected assembly program, a student teacher joining a class in January, a failure of the tape recorder, school closed because of bad weather, et cetera. In only two instances was it impossible to observe in consecutive days as a consequence of these uncontrollable conditions.

In all cases, the first day did not count (unbeknown to the teacher). This served as a means of acclimating the teacher and class to the observer's presence in the classroom and also allowed the observer to orient himself to the particular classroom situation. For example, the best place to sit to get a good tape recording was determined.

The schedule of visits was arranged in advance of observations. In each case, a "reminder" was sent to the teacher a week before the observation dates.

Observation Process in the Classroom

No codings (aside from practice) were done in the classroom. All discourse was tape recorded and an appropriate running commentary was produced by the observer. This technique made it possible to effectively "capture" the teaching that transpired. The rationale was that maintenance of discourse data would facilitate (1) accuracy of coding by allowing re-examination of discourse portions, (2) inter- and intra-reliability checks, and (3) the provision of a store of useful data which could be used for further analyses at a future time (meeting Objective Five).

During the observations, appropriate notations concerning details of incidents which might be difficult to categorize using only the tape recording were written down (running commentary). A short-hand notation was used. For example, "Bill had his hand up" might be necessary to later distinguish an "8" from a "9"; or "teacher frowned" could imply a "7" instead of a "3"; and so forth. Also included was a record of everything written on the chalk board by the teacher and students. It is important to note that this kind of communication was considered as verbal communication and was later categorized accordingly. Non-verbal communicative acts, such as head shaking, nodding and the like, were all noted. These, too, were eventually categorized as part of the communication.

The tape recorder used was a Webcor, EP2404 fully-transistorized portable. Both battery power and a power converter (transformer) were available. Recordings were made on $3\frac{1}{4}$ inch reels of 0.5 mil mylar tape at a speed of 1.88 inches per second. The entire instrument is no larger than a book and fitted easily within the observer's brief-case. Each tape, about 600 feet long, was sufficient for at least 100 minutes of time, providing both sides were used. With fifty minutes on a side, there was nothing lost because of tape switching.

The portable recorder had a number of advantages. The prime asset was that the observer could walk into the classroom, sit down, press a button, and immediately begin recording. There was no unpacking and tape threading, nor extension cords and other paraphernalia to distract the teacher and class. Also, although the teacher and many students knew that a recording was being made, they were not continually reminded of it as would be the case with a non-obscure standard machine. The only thing exposed was the microphone.

Usually the observer sat in the rear of the room or to one side. By placing the microphone on a soft cloth, it could be easily turned towards one speaker or another without introducing extra noise.

The observer became skilled in guessing when a comment (usually student's) would be inaudible on tape. In these cases the statement was written into the running commentary. Time notations were always written down so that, when listening to the tape in his laboratory, the observer could always readily synchronize the recording and the running commentary. In addition, the coding rate could be easily checked.

If the teacher was not communicating to the entire class, but instead walked about the room and talked to individual students or groups of students, as in laboratory and "seat work" situations, then the observer followed him about and recorded only individual confrontations. The recorder was held in one hand at the side as one would hold a book. The microphone, with its remote on-off switch, was in the other hand. During teacher-student confrontations, it was generally obscured. In time the observer became skilled in guessing the right volume control settings to insure good recordings.

For all activities, it is reasonable to say that with the techniques used, all teacher-student and student-student communication, not deliberately omitted (certain routines) was recorded. In many cases, acoustics in the room were bad. In the laboratory the din of student-student conversation was also a problem. On the whole, good audible tapes were produced throughout.

Finally, it is pointed out that teachers agreed to not give a test longer than ten minutes, nor show a long movie, nor take a field trip during the scheduled observations. Otherwise, they were to do whatever is ordinarily done (see Appendix B, Form B-3). Through the extensive discussions in the solicitation meetings and "spot check inquiries" of teachers, the investigator was assured that teachers made no special provisions concerning the observations, i.e., they in fact did "whatever they ordinarily do."

The Effect of the Observer's Presence in the Classroom

Six of nine items on a Teacher Questionnaire (see Appendix C) were developed to assess the effect of the observer's presence in the classroom as seen by the teacher. The questionnaire was mailed to teachers in early June, following completion of all observations. All were returned.

Each item concerning observer interference presented four choices from "very much" to "not at all." The items were scored on a basis of 1-2-3-4, with the high score in the desirable direction. The following list gives the item and the mean of seventeen teachers. A most favorable mean would be 4.0; the least favorable is 1.0.

Selected Items on "Teacher Questionnaire"

1. After my first visit, how distracted were you with my presence in the classroom?
   Mean = 3.4 (slightly).
2. After my first visit, how distracted was your class with my presence in the classroom?
   Mean = 3.1 (slightly).
3. To what extent, when I was there, and after the first visit, did your class refrain from participating in class discussions because of my presence in the classroom?
   Mean = 3.4 (slightly).
4. To what extent, when I was there, are after the first visit, do you think you taught differently from the way you usually teach?
   Mean = 3.6 (not at all).

5a. To what extent, after th first visit, were you hindered by an awareness of the fact that your classroom discourse was being recorded on tape?
Mean = 3.6 (not at all).

5b. To what extent, after the first visit, were you hindered by not being able to give a test (or show a film) of duration greater than ten or fifteen minutes?
Mean = 3.1 (slightly)

Table 5 summarizes the results for individual teachers. Comments were limited to expressing some inconvenience in carrying out the testing program.

TABLE 5

SCORES ON TEACHER QUESTIONNAIRE

Teacher	Item 1	2	3	4	5a	5b	6a	6b	7	Tot. Points Per Teacher[a]
01	3	4	4	4	4	4	4	3	3	33
02	3	3	3	4	3	4	4	3	3	30
03	3	3	3	3	2	2	3	3	3	25
04	3	2	2	4	4	4	3	2	3	27
05	4	3	4	4	4	4	3	3	3	32
06	4	3	4	4	4	2	3	2	3	29
07	3	3	4	4	4	4	4	3	4	33
08	4	3	4	3	4	4	4	3	4	33
09	3	3	4	3	4	3	4	4	4	32
10	4	4	4	4	4	3	3	3	4	33
11	3	3	2	3	4	2	4	3	3	27
12	3	4	4	4	3	3	4	3	3	31
13	3	3	4	4	4	3	3	3	3	30
14	4	3	4	4	4	4	2	2	4	31
15	4	3	3	4	3	4	4	3	4	32
16	3	2	2	3	3	3	2	3	3	24
17	4	4	3	3	4	4	4	4	3	33
Tot.[b]	58	53	58	62	62	57	58	50	57	515

[a] Total possible points per teacher on 9 items: maximum - 36; minimum - 9.

[b] Total possible points per item for 17 teachers: maximum - 68; minimum - 17.

In early March 1965 the sixty-item Student Questionnaire (SQ) on student attitude was administered. Four additional items inquiring how students perceived observer interference were attached (Appendix C). These provided the same four alternatives and were scored the same way as the above. Table 6 gives class mean scores for each item. Item 1 (61 on SQ) corresponds to Item 1 on Teacher Questionnaire. Similarly, 3 corresponds to 3 and 4 with 4. Item 2 (on how much the overall classroom situation was different from usual) has no direct correspondent.

It is clear that responses on both instruments were very favorable. It was concluded that observer interference was negligible.

Major Activities in the Classroom

The classroom discourse was divided into four Major Activities. The rationale behind this segmentation of "total teaching" is connected with a view that the teacher's influence may be different in various phases (activities) of the classroom process. Hence, it was necessary to produce Flanders matrices for each of the Major Activities as well as Total Teaching.

The Major Activities are mutually exclusive but are not inclusive of all activities in the classroom in view of certain omissions (e.g., routines). They were defined in the course of observations. It was necessary to determine

TABLE 6

STUDENT SCORES ON OBSERVER INTERFERENCE ITEMS

Class of Teacher	Class Means on Items[a]				Class Total		% Class Response
	1	2	3	4	$\bar{X}^b$	SD	
01	3.7	3.1	3.9	3.2	13.9	2.2	85
02	3.4	3.2	3.6	3.5	13.8	1.9	95
03	3.6	3.4	3.8	3.8	14.7	1.5	63
04	3.6	3.1	3.7	3.7	14.2	1.8	90
05	3.2	3.7	3.7	3.9	14.5	1.8	79
06	3.8	3.8	3.8	3.8	15.2	1.1	67
07	3.8	2.9	3.6	2.8	13.0	2.1	67
08	3.4	3.4	3.8	3.7	14.4	1.7	70
09	3.4	3.2	3.5	3.6	13.7	2.2	74
10	3.6	3.7	3.7	3.4	14.3	1.9	50
11	3.6	3.5	3.8	3.8	14.7	1.4	64
12	3.4	3.8	3.8	3.8	14.8	0.7	100
13	3.3	2.9	3.6	3.1	12.9	1.9	90
14	3.6	2.5	2.9	3.3	12.2	1.9	83
15	3.7	3.4	3.8	3.8	14.7	1.8	56
16	3.6	3.4	3.7	3.7	14.3	1.8	69
17	3.3	3.6	3.6	3.6	14.1	1.7	77

[a] Most favorable score per item, 4; least favorable, 1.

[b] Most favorable score, total of four items, 16; least favorable, 4.

through observation how the discourse could best be divided in line with the demands of the study. It is noted that codings from tapes were not done until all observations were completed.

The following list gives the final form of activity definitions as developed and used.

List of Major Activities

Activity 1. Planned Demonstration (DEM)

A teacher-planned demonstration, not a spontaneous illustration, but a demonstration obviously planned as a major activity of the lesson. This would generally be connected with concept development that is associated with new subject matter, although it could, partly or in whole, deal with concepts dealt with before.

Activity 2. Lecture (LEC)

Concept development through a teacher-planned oral presentation. Could involve audio-visual aids, but does not essentially involve laboratory apparatus as in DEM. Limited apparatus usage could come into it spontaneously. In essence, the teacher seeks to develop concepts and ideas in a planned constructive fashion. Short review to introduce the presentation is included.

Activity 3. Laboratory (LAB)

Students undertake laboratory experiment work. In this connection, the majority of the teacher's talk is directed to individuals and to small groups of students working together. Directions and announcements to the group, concerning lab, are included. Problem sessions in which the teacher discusses work with individuals at their seats are considered as "laboratory."

Activity 4. Recitation-Discussion (REC)

This is a discussion. It involves such things as going over homework, going over a test, an extended review of recent work, oral quizzing and oral drill. The teacher's "Any questions?" may be an invitation to this kind of activity. The discussion may be limited to particular items at hand as set by the teacher, or it may deal with topics put forth and developed, in large measure, by the students. Student questions and related discussion, if relatively extensive, would be included. So would extensive presentations of student ideas.

Note: Routines (giving assignment, taking attendance, etc.) are omitted when not part of a major activity. Aspects of some routines may be included. For example, a teacher's lecture presentation of material to provide background for an assignment is part of LEC; discussion of assignment may belong to REC even though the actual assigning is done following another activity.

In the above it is obvious that some review and discussion may be included in Lecture, some demonstration may be involved in Recitation-Discussion, and so forth. The essence of the activity was the deciding factor in categorizing the activities of the classroom. An open discussion activity, where student ideas predominate, was originally planned. There was so little of this found that it was necessary to fuse it with a closed discussion activity to form Activity 4 as defined above.

Coding the Discourse

Categorization of all classroom discourse was done in the investigator's laboratory following the completion of all observations. The coding of all tapes on one teacher were done before another teacher's were examined. Teachers were coded in the order in which they were observed in the fall. As the classroom was visited at least seven times (additional observations were made in cases of double periods and to get at least one laboratory on each teacher), it was an easy matter for the investigator to project himself into the class situation via the taped and written records (running commentary). It was just about as good as actually being there. It was better than live coding in the sense that the discourse could be stopped, started, and repeated at will. By doing all tapes on a teacher consecutively, full orientation to the "quirks" of a particular class was easier.

For each teacher, codings (Flanders category numbers) were separated according to activity and according to fall and spring. This was done to facilitate grouping of resultant matrices in various ways. For example, a teacher's fall DEM matrix was eventually compared to "spring DEM"; all REC of all teachers were ultimately combined to give a composite REC. The combination schemes used will be discussed in the next chapter. Thus with four activities, two series of observations, and seventeen teachers, a maximum of 136 separate sequences of numbers producing as many individual matrices were possible. (See Table 9, Chapter Six for a time-wise breakdown of activities.) Hence, activities were categorized as well as behavior.

Category definitions and ground rules employed were those described earlier. Numbers were recorded directly on Cornell Computing Center coding forms so that the data were set up for key-punching in one operation. Two "O's" were put between the last datum of a particular sequence and the first of the next quantity of data for a particular activity and series (fall-spring). Thus, every time there was a shift of activity, two "0's" were appended to foregoing data.

As expected, the running commentary was very necessary for accurate codings from tapes. A stop watch assisted timing and tape-commentary synchronization.

As a final point of interest, it was found that by "drawing" code numbers, the suggested rate of twenty categorizations per minute could be accurately maintained (within 5%).

Observer Reliability

An observer is "reliable" in categorizing behavior if he consistently produces the same sequence of numbers (categories) for a particular quantity of discourse. Two observers are "reliable" if they both produce the same data for the same observation. The latter is in some sense an indication of "validity," for if the two agree, it is likely they are both categorizing correctly. However, they could both be "wrong." Measures of reliability have been developed for the Flanders system; there are no direct measures of validity.

The measure used here was Scott's coefficient of reliability, Pi (Flanders, 1964a, p. 10). This reliability estimate is given by the relation:

$$\pi = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the proportion of category agreement and P_e is the proportion of agreement expected by chance. P_e is a function of P_i, the proportion of tallies in each category. For ten categories:

$$P_e = \sum_{i=1}^{10} P_i^2$$

In comparing two matrices, P_o is determined by subtracting the sum of absolute differences of category proportions from 1. That is, for ten-category matrices, A and B:

$$P_o = 1 - \sum_{i=1}^{10} \left| P_{i_A} - P_{i_B} \right|$$

As Flanders puts it, Scott's Pi gives "the amount that two observers exceeded chance agreement divided by the amount that perfect agreement exceeds chance." This calculation is one of several special modifications added by Wightman to his CDC 1604 Program for Interaction Analysis (Wightman, 1965). The program was used in performing the reliability calculations reported.

In all, 102 individual sets of data resulted in as many matrices. For each teacher a maximum of eight individual matrices were possible--four for the fall and four for the spring. In thirty-four instances, there was no observation of an activity for a teacher in one of the series of visits.

For each of the first three teachers coded, a twenty-minute sample was repeated immediately after the intial coding. In each case a reliability of .90 was found. After all codings were completed, ten individual matrices (of 102) were randomly selected for reliability checks. It was felt that reliability might vary with total number of tallies, type of activity, quality of tape recording, relative amount of teacher-student exchange, et cetera. Thus the purpose was to establish a figure for "overall reliability" of the matrices. Table 7 gives the results. Reliability extended from .93 to .65; the mean value was .81.

A Darwin Chi-Square Test was carried out for each pair of matrices compared. This statistic, a "likelihood-ratio criterion," tests the hypothesis that the frequency distributions in two or more matrices are the same (see Darwin, 1959; Flanders, 1964a, pp. 27-29). The test is discussed in Chapter Six. What is important here is that whereas the Scott Pi is sensitive only to row (or column) proportions, the Darwin statistic is a function of the distribution of tallies within a row or column. In nine of the ten comparisons, the hypothesis of no difference in distributions could not be rejected at the .10 level. In one case, it was rejected at the .01 level. Much of this

TABLE 7

INTRA-OBSERVER RELIABILITIES ON RANDOMLY SELECTED MATRICES

Matrix[a]	Time (min.)	No. of Tallies	X = I/(I+D)			Scott Rel.	Darwin Chi Sq.[b]	Sig. of Chi Sq.
			X	$\overline{X}$	SD			
0212a	7	99	.000	.000	.000	.648	8	NS[c]
0212b	--	127	.000					
0341a	37	692	.420	.466	.065	.751	38	NS
0341b	--	680	.512					
1141a	9	184	.191	.179	.018	.888	22	NS
1141b	--	221	.165					
0641a	26	523	.588	.610	.031	.855	40	NS
0641b	--	488	.631					
0241a	52	980	.253	.262	.013	.783	79	NS
0241b	--	911	.271					
0942a	26	525	.487	.521	.048	.826	26	NS
0942b	--	522	.555					
1322a	30	639	.299	.306	.010	.929	36	NS
1322b	--	633	.312					
0611a	11	202	.536	.528	.011	.743	33	NS
0611b	--	204	.520					
0142a	78	1539	.239	.255	.023	.850	122	$.010 < p < .025$
0142b	--	1440	.271					
0442a	44	903	.104	.116	.016	.827[d]	54	NS
0442b	--	881	.128					

[a] The first two digits of the matrix code give the teacher number; digit three is activity (1=demonstration, 2=lecture, 3=laboratory, and 4=recitation); for digit four, 1=fall and 2=spring; and a=original data, b=rerun for reliability check. Listing is in order in which reruns were made. All observations by Principal Investigator.

[b] df = 90.

[c] Not significant at the .10 level, i.e., $p > .10$.

[d] Mean Scott = .810; mean Chi Square = 34.

particular recording was relatively poor in quality. Mean chi-square was 34. Table 7 includes these results along with the "Scotts." It is noted that under the null hypothesis 99 per cent of the values in the sampling distribution of chi-square are greater than 62 when df equals 90.

As a further step, total I/(I+D) scores, "X" in Table 7, were compared. This score is a ratio of tallies in indirect influence categories (1..4) to tallies in all teacher categories (1..7). As this score was a major criterion for comparing teachers in terms of effectiveness, it was felt that an idea of the accuracy of it, too, should be obtained. The results (Table 7) show that a "rule of thumb" for this I-D percentage (so-called) is that a value found in a matrix in this study would have 68% confidence limits of approximately the value, plus and minus two units. In general, this would be on the order of the value, plus and minus 4 per cent of the value.

Two additional intra-observer reliability runs were carried out. A teacher whose tapes were typical in quality, amount of variation of activity, and in teacher-pupil exchange was "re-done" for the entire spring observation. A Scott of .89 was found; also, chi-square was not significant. This test indicated that Scott reliabilities of composite matrices of teachers, i.e., matrices of a teacher's Total Teaching for spring or fall, would be on the order of 0.9. It is likely that the typical reliability for a teacher's matrix based on all six observations would be a little higher.

For the above teacher, all of "spring laboratory" was compared with a re-run. Here Pi was .87 and chi-square was 40 (not significant). "Laboratory" was chosen as this is the most difficult activity for categorization purposes, other factors equal. The results comprise a portion of Table 8.

It should be noted that in all these tests, the point at which one activity ceased, on the tape, and another started had to be determined again in the re-run. This likely introduced additional error and lowered reliability values a little. Reliability measures with this included was desired, for the activity categorizations were involved in the data ultimately used.

A detailed analysis of the effects of variation of reliability, and of various ways of getting at reliability, was outside the scope of this study. However, the results shown here, facilitated a conclusion that the "overall reliability" was sufficiently satisfactory for the purposes of the project.

Finally, as a matter of interest, reliability checks were made with another observer who was also using the Flanders method in a study of teacher behavior in the practice teaching situation (Matthews, 1965). Two twenty-minute tape samples from this study and two from the other observer's were used in various comparisons. The inter-observer section

TABLE 8

INTRA AND INTER-OBSERVER RELIABILITIES ON NONRANDOMLY SELECTED MATRICES

Matrix[a]	Time (min.)	No. of Tallies	X = I/(I+D)			Scott Rel.	Darwin Chi Sq.	Sig. of Chi Sq.
			X	$\bar{X}$	SD			
Intra-Obs.								
12T2a	138	2790	.350	.329	.029	.891	87	NS[b]
12T2b	138	2719	.309					
1232a	33	678	.340	.348	.011	.871	40	NS
1232b	33	737	.356					
Inter-Obs.								
As	20	377	.322	.270	.074	.514	66	NS
Am	20	431	.217					
Bs	20	375	.310	.295	.022	.570	73	NS
Bm	20	377	.279					
Cs	20	321	.172	.127	.064	.309	60	NS
Cm	20	348	.081					
Ds	20	372	.399	.378	.030	.504	61	NS
Dm	20	376	.357					
(A+B)s	40	752	.317	.281	.051	.524	112	$.05 < p < .10$
(A+B)m	40	808	.245					
(C+D)s	40	693	.282	.247	.049	.463	97	NS
(C+D)m	40	724	.212					
(A+B+C+D)s	160	1445	.301	.265	.050	.493	185	$p < .005$
(A+B+C+D)m	160	1532	.230					

[a] Matrix 12T2 is all of teacher 12 in the spring; 1232 is teacher 12's spring lab; a=original data, b=rerun for reliability check. A, B, C, and D are selected teaching samples--A and B from teachers in this study, C and D from another study. Also, s denotes the Principal Investigator and m is the investigator who provided C and D.

[b] Not significant at the .10 level, i.e., $p > .10$. All comparisons are at df = 90.

of Table 8 shows the comparisons made and the detailed results. The reliabilities were typically on the order of 0.5, a value substantially less than intra-observer reliabilities in both studies.

Both observers are probably equally skilled in the basic Flanders technique. It is of interest to speculate on why the value was so low. For one thing, neither observer had visited classes of the other. It was clear that without having been at the scene when the recording was made, there was no way to mentally project one's self into the setting. This was probably the major factor. In addition, there was some variation of "special ground rules"; each observer used his own. Also, this investigator's experience had been in physics classrooms; the other's was in other science areas. This would have an effect on "8's" and "9's." Finally, an important factor may have been the lack of a written "running commentary" in these comparisons. These notions seem to point up the usefulness of the special techniques for observation developed in this study.

CHAPTER SIX

TEACHER-STUDENT VERBAL INTERACTION IN THE HIGH SCHOOL PHYSICS CLASSROOM

Introduction

Flanders categorization data give the amount, sequence and pattern of verbal behavior along a dimension of directness (or indirectness) of teacher influence. In this chapter, interaction matrices resulting from such data will serve as a basis for describing what takes place in the classrooms of Regents Physics teachers. This description meets Objective One of the study.

Several grand (composite) matrices are presented. This store of data, not only of current interest, may facilitate further analyses at a future time. In terms of a void of such data, it represents the fulfillment of Objective Five.

Two additional areas of concern are treated. First, the question of how well the distributions of numbers in the matrices represent certain consistencies of teacher behavior, i.e., teacher "style of teaching," is examined. Second, the results of an assessment of student attitude, and its connection with aspects of teacher influence, are made.

The Flanders Interaction Matrix

How the Matrix is Plotted

Flanders categorization data is a series of numbers, each of which represents the dominant form of behavior of a three-second time interval according to a specified system of categories and rules. The ten categories, the ground rules, and data collection techniques are described in Chapter Five.

The numbers are plotted pair-wise in a 10 X 10 matrix. In effect, the observer identifies a three-second interval of behavior according to one of 100 different matrix cells. This tally is made while the next interval is observed. A timing exception is that all category shifts within an interval are recorded.

The operation is done in two steps. Instead of tallying among 100 cells, the observer decides among ten categories. Consecutive pairs of numbers determine the cells. For example, a fifteen-second observation might produce the following numbers: 5,5,4,8,3. In plotting the matrix, one tally each would be placed, respectively, in the following matrix cells: 5-5, 5-4, 4-8, and 8-3. Cell x-y is the intersection of row x and column y in the matrix. Row and column designations correspond to categories, i.e., row 1 is associated with Category 1, column 3 is associated with Category 3, et cetera. The matrix is square. For n categories, there would be n^2 cells.

It is noted that all x-x cells (along the matrix diagonal) are "steady state," i.e., they do not involve any shift of category. On the other hand, x-y cells are "transition cells," i.e., they are tallied whenever a shift of category occurs from one interval to the next. Information is extracted from the matrix by examining the distribution of tallies among the 100 cells and by comparing tallies in various groups of cells.

Areas of the Matrix

Distributions of tallies in various groups of matrix cells (areas) provide quantitative measures concerning selected aspects of verbal interaction. Figure 1, a slight modification of a figure presented by Flanders (1963a, p. 7), will guide the interpretations that follow.

The figure indicates category definitions and outlines groups of cells that are generally of interest. Seven categories deal with teacher talk; of these, 1..4 are for indirect influence and 5..7 are for direct influence. Indirect influence fosters student participation (freedom of action); direct influence restricts it. Two categories deal with student talk (response and initiation). Category 10 is for both silence (pauses) and unclassifiable or non-identifiable communication (confusion).

Every time a particular category is used in the coding, a tally will result in one of the cells of the associated

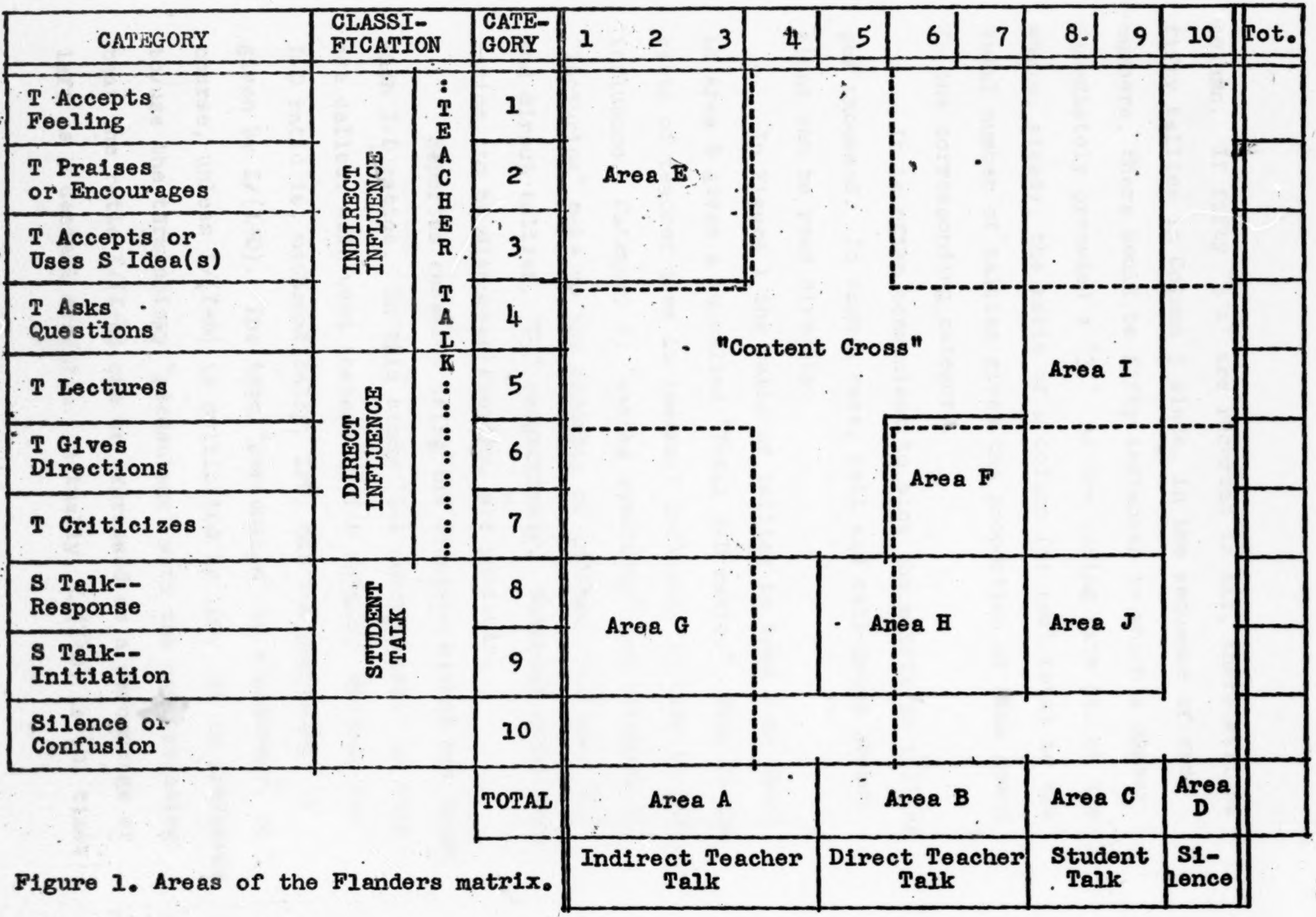


Figure 1. Areas of the Flanders matrix.

column. If fifty "3's" are recorded in all, there will be fifty tallies in Column 3 since, in the sequence of code numbers, there would be fifty instances in which a number immediately preceded a "3." As the coding rate is, on the whole, steady, the ratio of a column (or row) total to the total number of tallies gives the proportion of time spent in the corresponding category.

It is often convenient to plot the matrix in tallies per thousand. In such a case, cell and cell group proportions can be read directly.

In Figure 1 the ratio of tallies in Area A to those in Area B gives a so-called "Total I-D ratio." This is the ratio of teacher time in indirect influence to time in direct influence. Category 4, "asking questions" and Category 5, "lecturing" make up the majority of indirect tallies, "I," and direct tallies, "D," respectively. Numerous other such ratios, to be discussed shortly, are possible.

Reported research using the Flanders method has dealt with I-D ratios. In this study "I-D percentages," as they are called, were used instead of I-D ratios. Whereas the I-D ratio is, mathematically, I/D, the I-D percentage is given by I/(I+D). The term "percentage" is a misnomer, of course, unless I/(I+D) is multiplied by 100. It is preferred to use the terminology "percentage" with the understanding that the ratio I/(I+D) can be expressed as a percentage or left as a decimal fraction. Actually, I/(I+D) is .01 times per cent I. Nevertheless, the term "I-D percentage" seems most useful.

The distribution of an I-D ratio is generally skewed. An I-D percentage distribution will, on the other hand, tend to normality with an increasing number of values. As will be seen later, the standard deviation of the Total I-D percentage, where the latter is given by Area A divided by the sum of Areas A and B (Figure 1), is used as a measure of flexibility of influence. A distribution more nearly tending toward normality would be favored. In comparisons of indirect cells to direct cells, if there are very few direct tallies, a change of one tally in D can cause a large change in I/D. This is not the case if the ratio I/(I+D) is used. The ratio I/D theoretically extends from zero to infinity whereas I/(I+D) ranges from zero to 100. For these reasons it was decided in this study to deal with ratios of the form I/(I+D) instead of the form I/D. Transformations between I-D ratios and I-D percentages can always be readily carried out by the interested reader.

Returning to Figure 1, it is clear that the more indirect teacher will have a higher proportion of A + B in A.

The per cent of the time the teacher talks can be found through the ratio of A + B to total tallies, similarly for the per cent of student talk, C, and for silence or confusion, D. Also, it is sometimes of interest to know how teacher talk is distributed among the first seven categories.

Area E is a region of extended indirect influence. A teacher who shows awareness of student feeling, employs praise statements, and accepts and builds on student ideas will have relatively high tally loadings in this area. A concern with motivation through reward could be implied here as well. This is a characteristic of the more indirect teacher. Direct influence, on the other hand, would be associated with more tallies in Area F. This is a consequence of more directions, more criticism, and of shifts from directions to criticism and back again (6-7 and 7-6 cells). Loadings in the latter two cells can generally indicate of discipline problems. Heavy loadings in 6-7 and 7-6 in combination with high 6-9 loadings indicate resistance to teacher directions The ratio of E to E + F is called the "Extended I-D percentage."

Tallies in Areas G and H represent transitions from student talk to teacher talk. The distribution indicates how the teacher responds to student talk. An I-D percentage for these areas, the so-called "Rows 8 and 9 I/(I+D)," is given by G/(G+H). The more direct teacher is more likely to respond to student statements with his own ideas (8-5 and 9-5), or with criticism (8-7 and 9-7). The indirect teacher would have less of this and more acceptance and clarification praise, and encouragement as shown by heavier loadings in 8-2, 9-2, 8-3, and 9-3. The latter aspects of teacher indirectness in response are of a perfunctory nature, for teacher statements facilitating the tallies here are three seconds or less in extent. For consideration of developed acceptance, praise and clarification, the tallies in Area E would be examined.

Area I is composed of transition cells representing student talk following teacher talk (in Columns 8 and 9, Rows 1..7). Those in Column 8 are student response while Column 9 of Area I involves cells of student-initiated talk following teacher talk. An examination of the tally distribution in this area can provide indications of the kinds of teacher statements that foster student participation.

The amount of sustained student participation can be determined by the relative loading in Area J. These cells represent student talk for periods longer than three seconds. Extensive development of student ideas would, for example, produce a heavy loading in the 9-9 cell.

The comparison of tallies in the so-called "content cross" (area within dotted boundaries) to the remaining matrix gives an indication of the content orientation of the teaching observed. According to Flanders' studies, the median range of the teacher's questioning and provision of information is 75-80 per cent of teacher talk.

The foregoing serves to give an overview of the kind of information extractable from the Flanders matrix in terms of area comparisons. Individual I-D percentages and selected cells will be discussed in ensuing sections. The point to emphasize is that a great deal of information about what happened in the classroom can be gleaned from the plot. A reliable description of what took place from the standpoint of verbal interaction, on the directness dimension, is what the tool gives. The question of what should have happened, or of what is good and what is bad is something else. For reference purposes, Flanders' manual is suggested (1964a).

Selected I-D Percentages

A number of particular calculations involving ratios of tallies in groups of cells were of interest in this study. Some indications have already been given concerning the so-called "I-D percentages." The Total I/(I+D), Rows 8 and 9 I/(I+D), and Extended I/(I+D) were described in the preceding section.

Two additional "percentages" involving Row 8 and Row 9 are of interest. The Row 8 I-D percentage, given by the quotient (8-1 + 8-2 + 8-3)/(8-1 + 8-2 + 8-3 + 8-6 + 8-7), indicates the relative extent of positive motivation by the teacher (acceptance, praise, encouragement, etc.) immediately following student statements which are in the main, teacher-solicited. Note that teacher question and lecture statements, following termination of student statements in Category 8 are omitted in this "percentage." A similar calculation gives a so-called Row 9 I/(I+D) (same cells, but in Row 9). This I-D percentage indicates positive teacher motivation following student initiated statements. Flanders' research has shown that some teachers are accepting and motivating in response to teacher-solicited student statements only. That is, they are more direct in response to student initiated statements. This would show up as a "high" Row 8 I/(I+D) with a "low" Row 9 I/(I+D). At the junior high level, teachers who are equally accepting of both kinds of student statements appear to be more effective (Amidon and Flanders, 1963, p. 45).

A "Revised I-D percentage" indicates the kind of emphasis given to motivation and control. Given by i/(i+d) where i is the sum of Columns 1..3 and d is the sum of Columns 6 and 7, this I-D percentage compares indirectness to directness when the content aspects of the matrix (questions and lecturing) are removed. A more indirect approach to motivation and control will show up as a relatively higher value.

Emphasis on content can be found with the ratio of Columns 4 and 5 to the total matrix. This is a simple alternative to the "content cross" described earlier.

So-called "steady state" measures can be taken by examining various groups of matrix cells along the diagonal. All diagonal cells are non-transitional, i.e., a tally is entered in the x-x cell only when two x's are consecutively recorded. The relative amount of deliberation in the classroom is indicated by the ratio of the sum of diagonal cells to the total matrix. A steady-state I-D percentage, the quotient (1-1 + 2-2 + 3-3)/(1-1 + 2-2 + 3-3 + 6-6 + 7-7), would compare sustained indirect and direct influence.

Selected Matrix Columns, Cells, and Cell Groups

Particular cells and columns of the matrix may provide additional information concerning teacher influence in the classroom. The following outline summarizes characteristics of selected cells and columns in a comparison of direct and indirect teachers. The figures were reported by Amidon and Flanders (1963, pp. 38-49) as a result of normative data on junior high school teachers.

Cell and Column Directness-Indirectness Indicators For Junior High School Teachers

A. Column (Category) Indicators

	Mean % of Total for Teachers Who Are:	
Column	Direct	Indirect
1*	< 0.1	≥ 0.1
3*	2	9
4	8	11
6*	8	4
7	5	< 1
10	"more"	"less"

B. Selected Matrix Cells

Cell	For Direct Teacher	For Indirect Teacher
4-8	≥ 50% of Category 8	≤ 30% of Category 8
7-7	> 50% of Category 7	< 50% of Category 7
9-9	"low"	"high"
2-2	x	2x

* Means strong indicator.

Some comments on particular cells and cell loadings are in order. Only the more salient features will be described.

Flanders points out that extended praise, indicated by heavier loadings in the 2-2 cell, is praise which includes the reasons for teacher favor of student action. Although direct and indirect teachers (junior high school) have 2 per cent of their tallies in Category 2, indirect teachers exhibit twice as much extended praise (2-2 cell). Such praise is considered more effective. For one thing, it promotes student independence through teacher-goal clarification.

The 3-3 cell is one of particular importance. The teacher who makes a definite effort to employ student ideas and to use student suggestions in a discovery approach to concept development will have relatively higher loadings here coupled with "high" in 9-9. A deficiency of 3-3 tallies would imply little or no use of student ideas in the development of subject matter concepts. If student participation in the learning process is sought, this deficiency could signal a needed modification in interaction.

Thought provoking questions are indicated by heavier 4-10 loadings along with 10-8 and 10-9 tallies (especially the latter). But 4-10's with 10-4's would imply question repeats. On the other hand a drill situation would show few 10-8's and 10-9's and more 4-8 and 8-4 with less 4-10.

The 6-7 and 7-6 transitions have been cited as signals of discipline problems. Another indicator is the 5-7 cell. When the teacher shifts from lecture to criticism, an effort to maintain order during instruction is clear.

Loadings in the 6-6 cell would indicate extensive teacher directions. The proportion here may have bearing on the nature of a laboratory program (for example).

Student-student communication is indicated by the extent to which 9-10 and 10-9 loadings agree. Recall that 10's (i.e., 0's) are put between 9's when one student stops talking and another starts.

In the high school physics classroom an aspect of teacher behavior in the lecture activity that was either continually present or else non-existent, as the author went from class to class, was what he likes to refer to as "check pointing." By this is meant a teacher's hesitation preceded by a short question, usually an "Okay?" The teacher periodically ascertains whether or not students understand the development up to that point. He does not expect a response if all are "with him." The point is that the teacher provides an opportunity for students to "interrupt." It might also be assumed that such a teacher is alert to aspects of the subject matter that students have difficulty with, i.e., he is sensitive to the students. Interestingly enough, teachers either did this consistently in lecture or "never" did it. "Check pointing" would show up in terms of the extent to which 5-4, 4-10, and 10-5 tallies agree (teacher is lecturing, "checks if okay," goes on).

Another distribution of interest involves cells 4-8, 8-7, 7-2, 2-2, 2-8, and 8-3. The teacher who provides clues when the student fails to give information sought, and thus guides the student in discovering the correct response, might be distinguished by such a series of transitions. Note how the sequence goes. Teacher asks a question and rejects a wrong answer (4-8, 8-7). Teacher then gives information to clue the student (7-2, 2-2). Correct response results and is accepted (2-8, 8-3).

These combinations of several cells are simple examples of patterns of behavior for which reliable and quantitative information may be extracted from the matrix. A pattern associated with the teacher's building on student participation (the 3-3 cell was mentioned earlier) would involve loadings in 4-9, 9-9, 9-3, 3-3, and 3-5 or 3-4. Larger proportions would be in 9-9 and 3-3. The interpretation is that the teacher asks a "broad" question; student gives his ideas (x 9-9's); teacher accepts and uses student ideas, i.e., "builds" (y 3-3's) and teacher either asks another question or brings his own ideas into it. In physics classes some teachers rarely seek student ideas through participation. Detailed analyses of patterns in the matrix, beyond the scope of this report, suggest many challenging questions and possibilities concerning teacher-student interaction. Some early efforts in this regard are being made (Flanders, 1964c).

In this chapter normative data on physics teaching will facilitate some general descriptions of what takes place in the classroom. The foregoing sections show various ways in which reliable descriptive information can be gleaned from the interaction matrix. Subsequent chapters will deal with some aspects of the matrix and achievement of physics teachers. Comparisons will be made to the reported results concerning junior high school teachers. The following section is prefatory to the presentation of interaction data.

The Wightman Program for Interaction Analysis

All matrix plots and calculations in this study were performed by the Control Data 1604 computer at Cornell University. A program developed and specially modified by Wightman (1965) in line with the demands of this study, was used. Interaction (categorization) data was punched at seventy-four tallies per card.

The program plots the matrix with both actual tallies and tallies per thousand (i.e., per cent times ten) entered in the cells. With such a scheme, pairs of matrices are more easily compared by direct inspection. It is noted that one arrives at the same ratio regardless of which cell entry is used in a particular calculation.

The program handles matrices in sets. For each matrix of a set, the matrix is plotted and the following are listed: per cent each column is of the total matrix, per cent each teacher category is of teacher talk, per cent teacher talk, per cent content, and per cent steady state. In addition, the various I-D percentages, described earlier, are given. An option is included so that I-D ratios, or I-D percentages, or both may be obtained.

As many as ninety-nine matrices may comprise a single set; no more than ninety-nine sets may make up a "run." A "loop" is programmed to facilitate an "unlimited" number of consecutive runs with a single data "feed-in."

Following the completion of each set of matrices, the program puts all the data of the set together and calculates a composite or "summed" matrix. All calculations of individual matrices are repeated for the summed matrices. In addition, for each summed matrix, the percentage (or ratio) values are tabulated and means, variances, and standard deviations are given.

After all sets of a run are complete, a "grand summed matrix" is plotted; the set of percentage calculations for this matrix is carried out as well. A tabulation of Total I/(I+D) from each summed matrix is made and the mean, variance, and standard deviation is determined. A listing of total tallies from each of the summed matrices is included so that one can see at a glance the basis of each of the scores. (The more tallies "behind" a score, the more reliable the score.)

If sets are composed of two matrices, an option can be used providing Scott reliability determinations. If more than two, the Scott Pi would not be meaningful.

Finally, a Darwin Chi-Square test (to be discussed shortly) is carried out for each set of matrices. This test compares matrices in terms of proportions of tallies in the cells. If two matrices are involved, the statistic provides an indication of whether or not the two are alike; if more than two, the interpretation is one of homogeneity of matrices. The same test is repeated for the grand summed matrix in which the matrices compared are the summed matrices of sets in the run.

It is noted that the chi-square distribution closely approximates a normal distribution (with mean zero and standard deviation of 1) if the degrees of freedom exceed 90. In such cases the program provides for a transformation to a "z score," i.e., the number of standard deviations that the statistic is departed from the mean in the sampling distribution. For all Darwin calculations the program prints out the degrees of freedom and the significance of the result. Options are built in so that the user can elect to have or not have the Darwin calculation for any run.

I-D ratio and percentage claculations can conceivably involve a zero in the denominator. In such cases the program provides for the substitution of one for zero. A note to this effect is printed at the side of the matrix.

Computer time usage is surprisingly conservative. A series of four runs (involving three loops) with seventeen sets of two matrices in each run would involve a total of 136 individual matrices, 68 summed matrices, and 4 grand summed matrices. If all possible options were used, there would be sixty-eight Scott reliability determinations, sixty-eight Darwin tests of matrix pairs and four Darwin tests for homogeneity. In addition, 202 tables of various percentage values would be printed and 1,768 calculations each of the mean, variance and standard deviation would be done. Such a series would consume approximately ten minutes of computer time. A Darwin calculation for one pair of matrices might take as long as eight hours with a desk calculator. In this study four series of runs were required.

Interaction Matrices of High School Physics Teaching

Time in Major Activities

As mentioned earlier, individual matrices were based on a single teacher in one of the four Major Activities for one of the series of observations, i.e., fall or spring. These matrices were organized in various set arrangements in the series of runs. In a number of cases teachers did not undertake one or more of the activities during the fall or spring observations. Table 9 gives the distributions of observed interaction time among the Major Activities.

Each teacher was observed in a particular physics class for four consecutive days in the fall and for three consecutive days in the spring. The first visit was for

TABLE 9

DISTRIBUTION OF INTERACTION TIME AMONG MAJOR ACTIVITIES IN MINUTES

Tea-	1-DEMONSTRATION				2-LECTURE				3-LABORATORY			4-RECITATION				TOTAL-ALL ACTIVITIES			
cher	Fall	Spg	Tot	%[a]	Fall	Spg	Tot	%[a]	Fall	Spg	Tot	Fall	Spg	Tot	%[a]	Fall	Spg	Year	Year[b]
01	0	4	4	2	67	0	67	33	0	5	5	56	78	134	65	123	87	210	205
02	7	7	14	6	78	60	138	58	60	96	156	52	33	85	36	197	196	393	237
03	0	4	4	2	55	85	140	73	11	18	29	37	11	48	25	103	118	221	192
04	0	59	59	27	37	0	37	17	0	37	37	81	44	125	57	118	140	258	221
05	0	0	0	0	37	33	70	32	0	0	0	80	72	152	68	117	105	222	222
06	11	0	11	6	56	76	132	72	0	35	35	26	13	39	21	93	124	217	182
07	8	81	89	37	0	18	18	7	0	19	19	117	19	136	56	125	137	262	243
08	32	0	32	10	48	48	96	31	0	22	22	70	115	185	59	150	185	335	313
09	0	7	7	4	59	78	137	69	0	24	24	28	26	54	27	87	135	222	198
10	41	30	71	33	57	83	140	65	0	0	0	3	0	3	1	101	113	214	214
11	20	52	72	24	113	54	167	56	17	0	17	9	52	61	20	159	158	317	300
12	0	37	37	17	38	57	95	43	0	33	33	76	11	87	40	114	138	252	219

(continued)

TABLE 9 (continued)

Tea-cher	1-DEMONSTRATION				2-LECTURE				3-LABORATORY			4-RECITATION				TOTAL-ALL ACTIVITIES			
	Fall	Spg	Tot	%[a]	Fall	Spg	Tot	%[a]	Fall	Spg	Tot	Fall	Spg	Tot	%[a]	Fall	Spg	Year	Year[b]
13	0	0	0	0	30	30	60	30	0	20	20	74	63	137	70	104	113	217	197
14	15	35	50	26	48	85	133	69	37	37	74	11	0	11	6	111	157	268	194
15	28	7	35	20	15	28	43	25	0	54	54	61	33	94	55	104	122	226	172
16	4	37	41	18	41	41	82	37	24	0	24	67	33	100	45	136	111	247	223
17	93	65	158	72	23	37	60	28	0	30	30	0	0	0	0	116	132	248	218
TOTAL	259	425	684		802	813	1615		149	430	579	848	603	1451		2058	2271	4329	3750

[a] Percent this activity is of total for year excluding Activity 3 - Laboratory.

[b] Total for year, all activities excluding Activity 3 - Laboratory.

acclimation purposes only. In addition, an effort was made to visit at least one corresponding laboratory for each teacher. In some cases laboratory was scheduled at a time different from the class period.

Table 9 shows no laboratory time for Teachers 05 and 10. In the first case an unavoidable conflict arose and it was not possible to arrange a laboratory visit. Teacher 10 did not teach the laboratory program; this was done by another teacher.

In some cases, where double periods were involved in the schools, the observer, with the consent of the teacher, stayed on for the additional periods of observation. Figure A-5 (Appendix A) shows the variation of periods of instruction time in the schools.

The Grand Matrix of Regents Physics Teaching

For each teacher, a summed interaction matrix based on total observations in all Major Activities was plotted. These matrices are given in Appendix D as Tables D-1 through D-17. As is the case with all matrices presented in this report, both the actual number of tallies and the number of tallies per thousand are entered in the matrix cells. It is noted that in some cases the sum of the latter figure in the columns does not equal the total reported. This is due to rounding. Every entry of actual tallies was divided by the number of thousands of total tallies to get tallies per thousand.

Table 10 is the grand interaction matrix based on all observations of the seventeen Regents Physics teachers in all Major Activities. This matrix is a composite of all observations in the study. It contains over 86,000 tallies and represents more than seventy-two hours of teacher-student verbal interaction in the high school physics classroom. As such, and in view of the design of the study, it is properly viewed as normative data on physics teaching in central New York State at the present time. It is somewhat reasonable to view it as representative of high school physics teaching in general. In interpreting the grand matrix, it is sensible to view it as the matrix of the typical Regents Physics teacher.

As can be seen from the matrix, teacher talk and student talk are in proportion of 4.6 to 1. Silence (pauses) and confusion account for only 11 per cent of the time. Flanders has found that junior high school teachers provide about two-thirds of the classroom talk and that two-thirds of this talk is associated with direct influence. It is clear that physics teachers do considerably more talking (25% more). Of teacher talk, 73 per cent is in Columns 5, 6, and 7; hence they are also more direct. However, 92 per cent of teacher directness is due to Category 5, lecturing. This category accounts for 68 per cent of all teacher statements.

Of all teacher talk, 16 per cent involves questioning (Category 4). About half as often (8.5%), teacher statements

TABLE 10

GRAND INTERACTION MATRIX FOR ALL OBSERVATIONS OF 17 PHYSICS TEACHERS IN ALL MAJOR ACTIVITIES[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	83 0	5 0	3 0	25 0	53 1	3 0	1 0	2 0	13 0	13 0
2) T Praises or Encourages	5 0	241 3	27 0	84 1	99 1	3 0	2 0	154 2	188 2	122 1
3) T Accepts or Uses S Idea(s)	9 0	68 1	1843 21	983 11	1217 14	44 1	28 0	64 1	774 9	311 4
4) T Asks Questions	9 0	78 1	15 0	3572 41	322 4	28 0	22 0	2542 29	1648 19	2004 23
5) T Lectures	48 1	89 1	90 1	2509 29	34732 403	383 4	129 1	85 1	2121 25	2129 25
6) T Gives Directions	3 0	5 0	5 0	86 1	287 3	1164 14	28 0	32 0	161 2	417 5
7) T Criticizes	0 0	10 0	10 0	97 1	285 3	38 0	473 5	49 1	274 3	116 1
8) S Talk--Response	3 0	66 1	1416 16	678 8	660 8	42 0	146 2	384 4	81 1	198 2
9) S Talk--Initiation	18 0	165 2	1819 21	877 10	2244 26	159 2	426 5	5 0	3479 40	1230 14
10) Silence or Confusion	23 0	198 2	113 1	1329 15	2416 28	324 4	97 1	357 4	1683 20	3018 35
Total	201	925	5341	10240	42315	2188	1352	3674	10422	9558
Per Cent Times 10	2	11	62	119	491	25	16	43	121	111

---

[a] Total tallies 86,216--based on 4,329 minutes (72 hours) of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

are associated with acceptance and clarification. Thus, considerably less than 10 per cent of teacher talk is associated with concept development in which the teacher builds on student ideas and suggestions.

About 1.5 per cent of teacher statements involve praise and encouragement. Less than 0.3 per cent of teacher talk appears in Category 1, acceptance of feeling. Thus, the employment of social skills connected with aspects of positive motivation is rare in the high school physics classroom.

In contrast there is a somewhat larger proportion of teacher talk devoted to giving directions and criticism (3.4% and 2.2%, respectively). However, as will be seen later, this is substantially less than what is reported for junior high school teachers.

Thus the examination of teacher talk in the physics classroom shows the teaching to be essentially "telling." The implication of relatively little learning through discovery may come as a surprise to both educational psychologists and curriculum builders. However, it is emphasized that the matrix interpretations in this chapter are limited to a description of "what is." They do not deal with "what ought to be."

By comparing Areas A and B (Figure 1) using data of Table 10, it is seen that the Total I/(I+D) for the typical Regents Physics teacher is 27 per cent. If one eliminates Categories 4 and 5, a Revised I-D percentage of 65 per cent results. Thus, when the teacher is not lecturing and questioning, he is providing indirect influence more often than direct influence. This implies that the emphasis given to motivation and control is more on the "indirect side," i.e., somewhat more positively affective.

The matrix shows an Extended I-D percentage of 58 per cent. This indicates slightly more sustained usage of Categories 1, 2, and 3 than sustained 6 and 7. This, too, supports the notion that when content emphasis is removed (Categories 4 and 5), the physics teacher is "slightly indirect" in his influence. However, it is noted again that content emphasis accounts for 84 per cent of teacher talk.

Tallies in Rows 8 and 9 of Table 10 may be examined to ascertain teacher response to student talk. The grand matrix shows a Rows 8 and 9 I-D percentage of 57 per cent. Thus, the physics teacher is a bit more likely to respond to student statements with indirect influence than with direct behavior. This means a little more acceptance, clarification, encouragement, et cetera than his own ideas, directions and criticism.

Of particular interest are Row 8 I-D and Row 9 I-D percentages of 90 and 77 per cent respectively. It is clear that the physics teacher is substantially more accepting of student statements that he has solicited (Row 8) than he is of student initiated statements. This is a striking result in view of a marked lack of student misbehavior (socially) in high school physics classrooms. It supports the implication of a failure to build on student ideas.

An examination of individual cells will be enhanced with reference to the outline given earlier. The 4-8 cell represents more than 67 per cent of Category 8 while the 7-7 cell is 31 per cent of Category 7. The first indication is one of direct influence while the low 7-7 loading implies indirectness. This discrepancy with Flanders may be rationalized in terms of the substantially greater maturity of physics students as opposed to youngsters in junior high school. The need for normative data at the senior high school level is illustrated by such a disparity.

The 2-2 cell contains 0.3 per cent of all tallies. This shows that extended praise and encouragement are rarely used. Flanders' teachers used even less; however, the special ground rule concerning subject matter clueing likely put some tallies in this cell that Flanders may have put elsewhere.

It is noted that the 9-9 loading is ten times greater than that of 8-8 while the ratio of Column 9 to Column 8 is 2.8. This arises from the relatively short responses comprising 8's while student initiated statements are generally much more extensive. The significant point for physics teaching, however, is that there are many more 9's than 8's while teachers are more accepting of 8's than 9's. The nature of what is dealt with in the classroom (the physics) is probably responsible for the high 9 to 8 ratio. Thus, there is more opportunity for the teacher to build on student ideas, i.e., more chance for discovery. Yet, as already mentioned, teachers do little of it; they prefer to give their own ideas and opinions, i.e., to "tell." The 9-9 to 3-3 comparison of 2 to 1 (4% and 2%, respectively) adds support to this conclusion.

The 4-10 and 10-9 loadings are similar (2.3% and 2.0%). But only 0.4 per cent of the tallies are in 10-8. This indicates some amount of thought provoking questions.

The relative void of 6-7, 7-6, and 5-7 tallies substantiates an earlier comment concerning good social behavior of students in physics classes. Discipline is not a problem in these classes.

The notion of teacher "check pointing" was discussed earlier. This is a pattern in which the teacher provides opportunities for student interruptions during lecture. It was pointed out that an indication of this would be the extent to which 5-4, 4-10, and 10-5 loadings were similar. Table 10 gives 2.9, 2.3, and 2.8 per cent respectively. As will be seen shortly, this is much less common in the junior high situation.

In contrast, the pattern described for student clueing does not show up. The lack of student clueing goes hand in hand with the lack of use of student ideas, discussed earlier.

The major value of the grand matrix presented lies in its normative aspects. It will provide an average array to which an individual observation may be compared. Some of the interpretations in the foregoing may suggest interesting questions about what changes in the average would be desirable and how such changes might be effected through science teacher education.

Grand Matrices of Major Activities

Tables 11, 12, 13, and 14 present the summed (grand) matrices for each of the Major Activities (Demonstration, Lecture, Laboratory, and Recitation-Discussion[1] respectively). These matrices, like the grand matrix of Table 10, are composites. They are composed of all observation data collected for the particular activity. The reader may refer to Table 9 for individual teacher contributions in terms of tallies.

Figures 2 and 3 show category comparisons of the five grand matrices. (The reader is alerted to the fact that Category 5 bars of Figure 3 extend way beyond the limits of the plot.)

Two features are at once obvious in Figure 2. There is substantially more of Category 5 (lecture) in the Lecture Activity and substantially less in Laboratory. The ratio is about 2 to 1. If only teacher talk is considered (Figure 3), the ratio is about 1.5 to 1.

---

[1] The terms "recitation" and "recitation-discussion" are used interchangeably as a matter of brevity.

TABLE 11

SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 15 PHYSICS TEACHERS IN ACTIVITY ONE - DEMONSTRATION[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	8 1	0 0	0 0	3 0	6 0	0 0	0 0	0 0	2 0	1 0
2) T Praises or Encourages	0 0	39 3	5 0	13 1	21 2	0 0	0 0	11 1	24 2	25 2
3) T Accepts or Uses S Idea(s)	1 0	7 1	274 20	150 11	181 13	3 0	1 0	6 0	130 10	67 5
4) T Asks Questions	0 0	9 1	3 0	489 36	49 4	3 0	1 0	290 22	301 22	304 23
5) T Lectures	7 1	17 1	21 2	351 26	5294 394	47 4	12 1	5 0	368 27	497 37
6) T Gives Directions	0 0	1 0	1 0	6 0	34 3	58 4	1 0	0 0	9 1	45 3
7) T Criticizes	0 0	0 0	1 0	8 1	30 2	1 0	26 2	2 0	28 2	12 1
8) S Talk--Response	1 0	3 0	178 13	67 5	60 4	1 0	11 1	13 1	4 0	27 2
9) S Talk--Initiation	2 0	33 2	310 23	146 11	386 29	14 1	47 4	0 0	463 34	286 21
10) Silence or Confusion	1 0	29 2	27 2	216 16	558 42	28 2	9 1	38 3	358 27	799 60
Total	20 1	138 10	820 61	1449 108	6619 493	155 12	108 8	365 27	1687 126	2063 154

[a] Total tallies 13,424--based on 684 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE 12

SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 17 PHYSICS TEACHERS IN ACTIVITY TWO - LECTURE[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	44 1	3 0	2 0	11 0	27 1	2 0	0 0	1 0	3 0	2 0
2) T Praises or Encourages	5 0	99 3	12 0	28 1	35 1	1 0	0 0	74 2	59 2	24 1
3) T Accepts or Uses S Idea(s)	4 0	22 1	757 23	334 10	503 16	5 0	7 0	18 1	239 7	74 2
4) T Asks Questions	3 0	34 1	5 0	584 49	117 4	2 0	1 0	785 24	377 12	708 22
5) T Lectures	25 1	31 1	35 1	943 29	17386 536	57 2	23 1	35 1	611 19	674 21
6) T Gives Directions	2 0	1 0	1 0	7 0	52 2	94 3	0 0	1 0	4 0	27 1
7) T Criticizes	0 0	4 0	2 0	22 1	80 2	1 0	96 3	11 0	36 1	13 0
8) S Talk--Response	1 0	27 1	533 16	196 6	224 7	1 0	34 1	79 2	10 0	36 1
9) S Talk--Initiation	2 0	45 1	590 18	145 4	575 18	6 0	86 3	0 0	985 30	186 6
10) Silence or Confusion	9 0	71 2	26 1	346 11	821 25	20 1	18 1	137 4	296 9	643 20
Total	95 3	337 10	1963 61	3616 111	19820 611	189 6	265 8	1141 35	2620 81	2387 74

[a] Total tallies 32,433--based on 1,615 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE 13

SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 15 PHYSICS TEACHERS IN ACTIVITY THREE - LABORATORY[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	3 0	1 0	0 0	2 0	4 0	0 0	0 0	0 0	2 0	1 0
2) T Praises or Encourages	0 0	13 1	1 0	8 1	12 1	1 0	0 0	7 1	18 2	21 2
3) T Accepts or Uses S Idea(s)	0 0	12 1	67 6	85 7	110 10	19 2	5 0	4 0	82 7	51 4
4) T Asks Questions	2 0	4 0	2 0	241 21	27 2	13 1	5 0	351 30	406 35	162 14
5) T Lectures	3 0	10 1	9 1	249 22	2434 211	169 15	23 2	7 1	437 38	392 34
6) T Gives Directions	0 0	2 0	2 0	39 3	128 11	787 68	11 1	2 0	113 10	256 22
7) T Criticizes	0 0	1 0	1 0	11 1	45 4	23 2	80 7	7 1	65 6	44 4
8) S Talk--Response	0 0	7 1	92 8	99 9	91 8	25 2	22 2	8 1	18 2	46 4
9) S Talk--Response	3 0	16 1	252 22	253 22	525 46	107 9	103 9	0 0	454 39	381 33
10) Silence or Confusion	2 0	15 1	9 1	226 20	357 31	196 17	28 2	22 2	499 43	575 50
Total	13 1	81 7	435 38	1213 105	3733 324	1340 116	277 24	408 35	2094 182	1929 167

[a] Total tallies 11,523--based on 579 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE 14

SUMMED INTERACTION MATRIX FOR ALL OBSERVATIONS OF 16 PHYSICS TEACHERS IN ACTIVITY FOUR - RECITATION[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	28 1	1 0	1 0	9 0	16 1	1 0	1 0	1 0	6 0	9 0
2) T Praises or Encourages	0 0	90 3	9 0	35 1	31 1	1 0	2 0	62 2	87 3	52 2
3) T Accepts of Uses S Idea(s)	4 0	27 1	745 26	414 14	423 15	17 1	15 1	36 1	323 11	119 4
4) T Asks Questions	4 0	31 1	5 0	1258 44	129 4	10 0	15 1	1116 39	564 20	830 29
5) T Lectures	13 0	31 1	25 1	966 33	9618 334	110 4	71 2	38 1	705 24	566 20
6) T Gives Directions	1 0	1 0	1 0	34 1	73 3	225 8	16 1	29 1	35 1	89 3
7) T Criticizes	0 0	5 0	6 0	56 2	130 5	13 0	271 9	29 1	145 5	47 2
8) S Talk--Response	1 0	29 1	613 21	316 11	285 10	15 1	79 3	284 10	49 2	89 3
9) S Talk--Initiation	11 0	71 2	667 23	333 12	758 26	32 1	190 7	5 0	1577 55	377 13
10) Silence or Confusion	11 0	83 3	51 2	541 19	680 24	80 3	42 1	160 6	530 18	1001 35
Total	73 3	369 13	2123 74	3962 137	12143 421	504 17	702 24	1760 61	4021 139	3179 110

[a] Total tallies 28,836--based on 1,451 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

K+E 10 X 10 TO THE CM. 358-14
KEUFFEL & ESSER CO. MADE IN U.S.A.

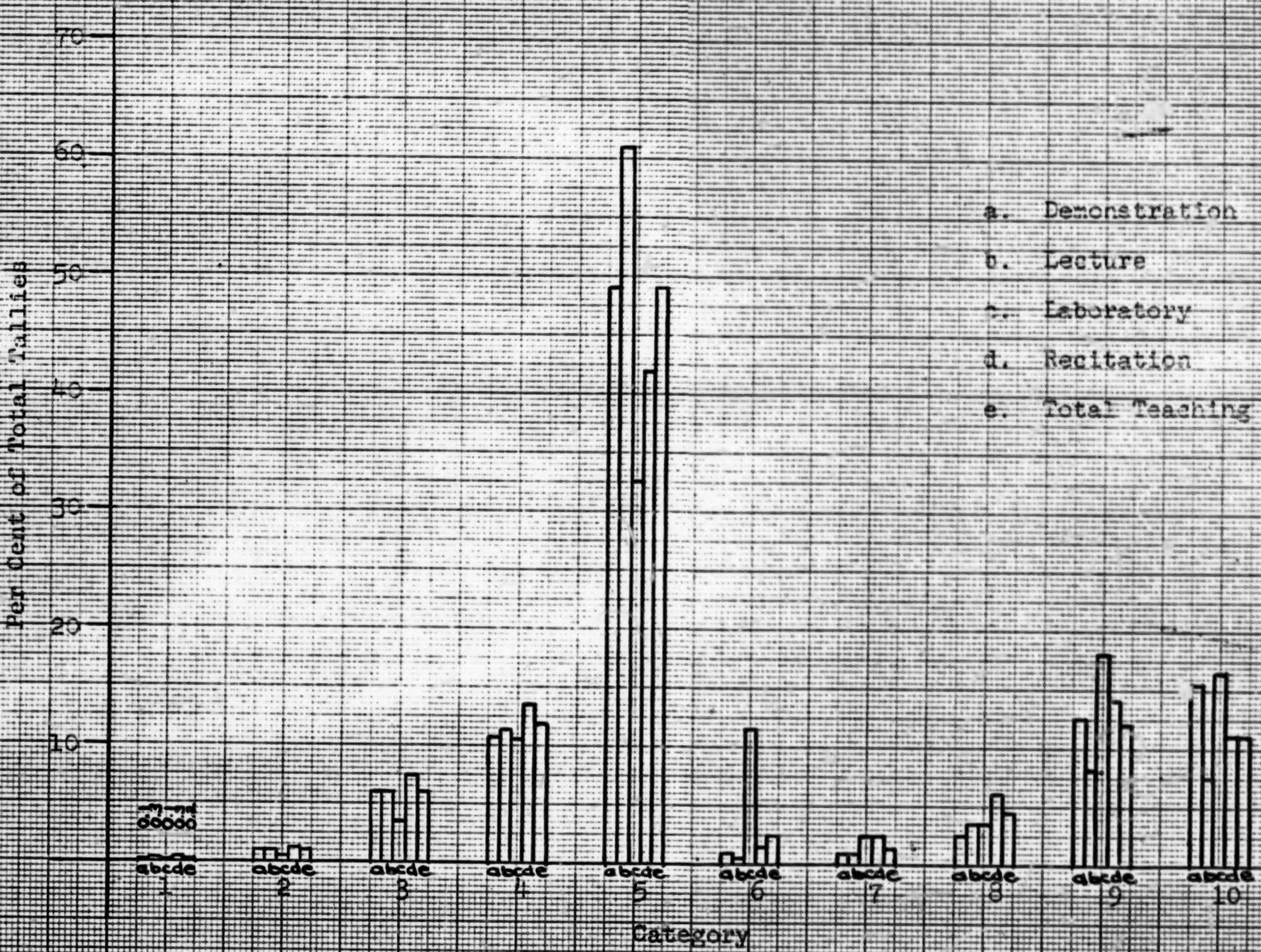


Figure 2. Distributions of tallies in Flanders categories

359-14 KEUFFEL & ESSER CO.
Millimeters, 5 mm. lines accented, cm. lines heavy.
MADE IN U.S.A.

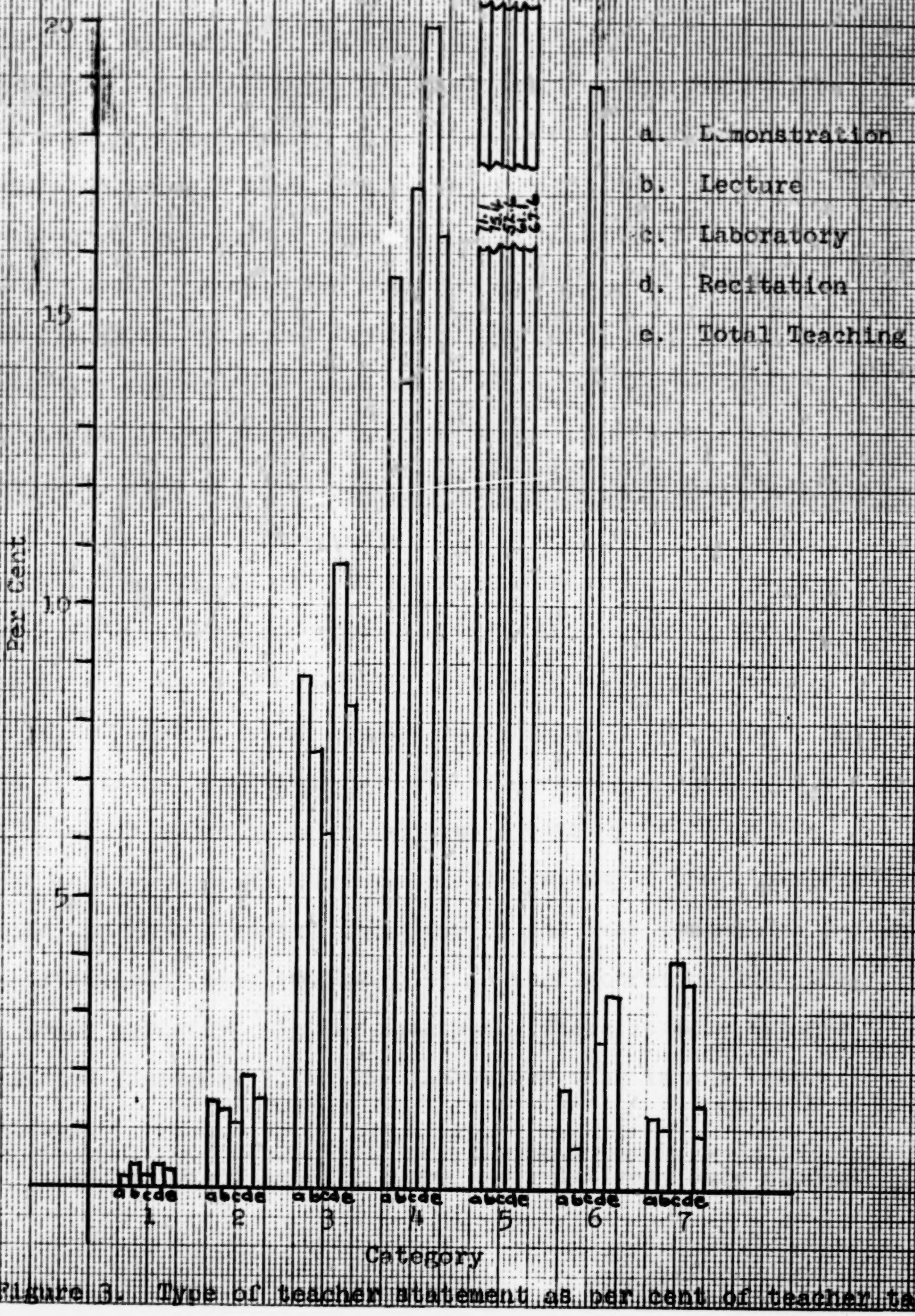


Figure 3. Type of teacher statement as per cent of teacher talk.

A striking feature of both figures is the severe increase of teacher directions (Category 6) in the Laboratory Activity. It should be recalled that the beginning of the laboratory period in which the teacher announces directions to the entire group makes up only a small amount of the data of this activity. Also, seat work, included in this activity, contributed less than 2 per cent. The observer followed the teacher about the room and recorded discourse as the teacher went from group to group and from student to student. One would thus expect that there would be a substantial amount of Category 9 and perhaps considerable encouragement and acceptance from the teacher. Physics curriculum developers and teachers alike would agree that if there is to be any student discovery at all, it would surely be in the laboratory. It is a common notion that in the laboratory the student will have an opportunity to do some physics on his own, i.e., that he will do some "sciencing." Yet 19 per cent of teacher statements in the laboratory are directions and commands to which students are to comply. This is six times more directions than any other activity. There is less praise and encouragement (Category 2) and less acceptance and use of student ideas (Category 3) in the laboratory than in any other activity. In the same vein, there is more teacher criticism (Category 7) in the laboratory than in any other activity.

As one might expect, there are relatively more student initiated statements in the laboratory than in the other activities (Figure 2). That the proportion of 9's is second highest for Recitation is reasonable. Even though Category 5 is lowest for Laboratory and Category 9 is highest in the same activity, there are still almost twice as many 5's as 9's in the laboratory.

Some features of the Demonstration Activity are interesting as well. Here the teacher is developing new concepts through a planned demonstration. One should expect some amount of student discovery here too, i.e., one should expect less teacher "telling." There is more. Figure 3 shows Category 5 to be second highest in this activity. Only in the lecture activity are there proportionately more 5's. There is only an average proportion of 9's. This suggests no increase in student questions. The relatively higher percentage of 10's in this activity is reasonable, however.

Finally, Figure 2 clearly shows the relatively small usage of statements fitting Categories 1, 2, and 7 (especially 1).

Further Comparisons of Interaction Matrices

Data from two grand matrices reported by Flanders (1964a, pp. 41-43) on eighth grade mathematics and seventh grade social studies were compiled with the grand matrix of physics teaching from this study. To facilitate easy comparisons, Table 15 was prepared. For this table the three sets of data were put into a single matrix using triple entries of tallies per thousand in the cells.

In the remarks concerning aspects of the grand matrix of physics teaching (Table 10), a number of comments involved comparisons to these data on junior high school teachers. A few additional points are clear in an examination of Table 15. Note that in each cell the top number is for physics teachers while the middle and bottom correspond to Flanders' data.

For one thing, even though this study reports only 0.3 per cent of all tallies in the 2-2 cell for physics teachers, this is three times as much as that reported for junior high school teachers. Also, physics teachers have over 25 per cent more 3-3 tallies.

There is considerably more sustained lecture in physics classes (40%) with less in eighth grade mathematics (34%) and least in seventh grade social studies (25%).

Comparisons of the 6-7, 7-6, and 5-7 cells show marked differences between physics and junior high. It is clear that the latter devote substantially more time to disciplining students.

Following solicited student statements, junior high teachers use much more praise and encouragement than is the case with physics teachers (8-2 cell). Physics teachers are more accepting of student initiated statements (9-3).

A striking disparity is seen in the 8-8 cell (sustained student solicited statement). Note that only 0.4 per cent of

TABLE 15

CELL-WISE COMPARISONS OF GRAND MATRICES OF HIGH SCHOOL PHYSICS, EIGHTH GRADE MATHEMATICS AND SEVENTH GRADE SOCIAL STUDIES TEACHERS[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	1[b]	0	0	0	1	0	0	0	0	0
	0[c]	0	0	0	1	0	0	0	0	0
	0[d]	0	0	0	0	0	0	0	0	0
2) T Praises or Encourages	0	3	0	1	1	0	0	2	2	1
	0	1	2	2	4	1	0	0	1	2
	0	1	2	3	3	1	0	0	1	1
3) T Accepts or Uses S Idea(s)	0	1	21	11	14	1	0	1	9	4
	0	1	16	10	14	1	0	2	4	3
	0	1	16	13	14	1	1	3	6	5
4) T Asks Questions	0	1	0	41	4	0	0	29	19	23
	0	0	0	16	5	2	2	70	5	8
	0	0	1	21	5	2	1	60	10	7
5) T Lectures	1	1	1	29	403	4	1	1	25	25
	1	1	1	40	343	14	5	2	14	12
	0	1	1	29	247	11	4	3	12	13
6) T Gives Directions	0	0	0	1	3	14	0	0	2	5
	0	0	0	3	8	20	2	13	4	13
	0	0	0	4	7	26	3	7	5	16
7) T Criticizes	0	0	0	1	3	0	5	1	3	1
	0	0	0	2	5	2	12	1	2	6
	0	0	0	4	4	3	15	1	2	5
8) S Talk--Response	0	1	16	8	8	1	2	4	1	2
	0	6	19	22	27	8	3	29	3	3
	0	5	20	19	18	5	3	115	5	5
9) S Talk--Initiation	0	2	21	10	26	2	5	0	40	14
	0	2	12	5	16	5	3	0	20	4
	0	3	19	6	12	5	2	0	40	5
10) Silence or Confusion	0	2	1	15	28	4	1	4	20	35
	0	1	0	8	12	10	5	3	12	62
	0	1	1	9	12	14	6	3	11	57
Totals	2	11	60	117	491	26	14	42	121	110
	1	12	50	108	435	63	32	120	65	113
	0	12	60	108	322	68	35	192	92	114

(continued)

TABLE 15 (continued)

---

[a] Grand matrices for eighth grade mathematics and seventh grade social studies teachers are those reported by Flanders (1964a, pp. 41-43).

[b] Cell entry for the grand matrix of Regents physics teachers. The number indicates tallies per thousand, i.e., tenths of one per cent of total tallies, and is based on a total of 86,216 tallies from approximately 71.8 hours of interaction in 17 classes of a central New York State region. Each class was observed at least six hours by the Principal Investigator.

[c] Cell entry for the grand matrix of eighth grade mathematics teachers. The number indicates tallies per thousand, i.e., tenths of one per cent of total tallies, and is based on a total of 58,614 tallies from approximately 48.8 hours of interaction in 16 classes of Minneapolis and St. Paul. "All classes were observed for a minimum of four hours, most at least six hours" (Flanders, 1964a, p. 39). Observations were performed by the research staff of Dr. Ned A. Flanders.

[d] Cell entry for the grand matrix of seventh grade social studies teachers. The number indicates tallies per thousand, i.e., tenths of one per cent of total tallies, and is based on a total of 51,835 tallies from approximately 43.2 hours of interaction in 16 classes of Minneapolis and St. Paul. "All classes were observed for a minimum of four hours, most at least six hours" (Flanders, 1964a, p. 39). Observations were performed by the research staff of Dr. Ned A. Flanders.

the tallies are in this cell for physics teachers but 2.9 per cent for eighth grade mathematics and 11.5 per cent for seventh grade social studies. The three are not so different in the 9-9 cell, but 8-8 to 9-9 comparisons show wide differences.

There are also extensive differences in usage of certain categories as indicated by column totals. Junior high teachers spend more than twice as much time giving directions (Category 6) and criticizing (Category 7). Much more of the student talk is in Category 8; the ratios of Column 9 to Column 8 are 3 for physics teachers and about 0.5 for the others.

Table 16 organizes selected cell and column percentages from all the matrices under discussion in this section. Table 17 does the same thing for the various I-D percentages of interest. It is pointed out that the grand matrices of physics Major Activities are based on much fewer total tallies (individually) than the grand matrix for "Total Teaching" of Regents Physics and the two grand matrices reported by Flanders. Although not as reliable, as a consequence, the activity matrices should be reasonable estimates nevertheless.

Although there is considerable variation within all the measures of both tables, only a few of the more salient differences will be mentioned here.

For seventh grade social studies, sustained student talk, given by diagonal 8 and 9, i.e., cells 8-8 and 9-9, is

TABLE 16

COMPARISONS OF SELECTED CELL AND COLUMN PERCENTAGES FROM COMPOSITE FLANDERS MATRICES

Composite Matrix	Tallies	% Teacher Talk is of Total Talk (1..7)/(1..9)	% Col. 5 is of Teacher Talk (5)/(1..7)	% of Total in 5-5 Cell	% in Diagonal 8 and 9[a]	% Content (4+5)/tot	% Steady State[b]
Demonstration	13,424	81.9	71.1	39.4	3.5	60.1	55.6
Lecture	32,433	87.5	75.4	53.6	3.2	72.3	67.1
Laboratory	11,523	73.9	52.6	21.1	4.0	42.9	40.5
Recitation	28,836	77.5	61.1	33.4	6.5	55.8	52.4
Total Teaching (Regents Phys.)	86,216	81.6	67.6	40.3	4.4	61.0	56.8
Eighth Grade Mathematics[c]	58,614	79.1	61.8	34.3	4.9	54.3	50.9
Seventh Grade Social Studies[c]	51,835	67.9	53.3	24.7	15.5	43.0	53.8

[a] Sum of 8-8 and 9-9 cells divided by total times 100.

[b] Sum of all cells along matrix diagonal divided by total times 100.

[c] Calculations made from grand matrices reported by Flanders (1964a, pp. 41-43).

greater by about a factor of three than in any of the others (Table 16).

Table 17 shows that the average Total I-D percentage for physics is between the two junior high figures. The variation of this measure among the activities in physics is considerable.

The revised I-D percentage shows marked differences. Excluding emphasis on content, physics teachers are more indirect on the whole (65%) compared to junior high teachers (about 40%). The striking decrease in the case of physics Laboratory Activity (Revised I-D = 24.6%) is a consequence of extensive teacher directions and direct control.

Only seventh grade social studies teachers are equally accepting of the two types of student statements. In all physics activities, the solicited student statements, i.e., responses to "narrow" questions are more readily accepted by teachers. That is, Row 8 I-D percentage is consistently greater than Row 9 I-D percentage.

Earlier comments about physics laboratory are supported by the values of the Extended I-D percentage (comparison of extended indirect influence to extended direct influence). The figures show that, not counting Laboratory, physics teachers make more use of extended indirect influence than do junior high teachers. The latter are more direct (i.e., more direction and criticism with less praise, encouragement, and acceptance). However, in the physics laboratory the

TABLE 17

COMPARISONS OF SELECTED I-D PERCENTAGES FROM COMPOSITE FLANDERS MATRICES

Composite Matrix	Tallies	Total I/(I+D)	Revised	Row 8	Row 9	Rows 8 & 9	Extended	Col.9/ Col.8
Demonstration	13,424	26.1	78.8	93.8	85.0	58.8	79.5	4.67
Lecture	32,433	22.9	84.1	94.1	87.4	62.4	83.2	2.31
Laboratory	11,523	24.6	24.6	67.8	56.3	45.3	9.7	5.20
Recitation	28,836	32.8	68.0	87.2	77.1	60.0	63.3	2.28
Total Teaching (Regents Phys.)	86,216	26.7	64.7	89.5	76.7	57.4	57.8	2.81
Eighth Grade Mathematics[a]	58,614	24.7	40.9	69.4	63.6	51.6	35.7	0.54
Seventh Grade Social Studies[a]	51,835	30.0	41.6	75.8	75.9	61.5	29.9	0.47

[a] Calculations made from grand matrices reported by Flanders (1964a, pp. 41-43).

teacher is much more direct on this measure. The change from 83.2 per cent in Lecture Activity to 9.7 per cent in Laboratory is extreme. Again, in terms of commonly accepted goals of laboratory instruction, one would expect this measure to be greater, if different. One would certainly not expect it to be less by a factor of eight.

As mentioned earlier, the purpose here was to set forth the data for the extensive systematic observations of physics teaching that were carried out. The discussion and comments concerning the individual matrices and the comparisons within physics and between physics and junior high school subjects is by no means exhaustive. The data are arranged to aid the interested reader in exploring further the various aspects of the matrices.

The Interaction Matrix as an Indicator of Teacher Style

Consistency of Matrices - A Rationale

The distribution of tallies in the Flanders matrix results from objective observations of teacher and student behavior on the dimension directness of teacher influence. Teachers likely have certain consistencies concerning the ways in which they interact with students in the classroom. Consistent aspects of teacher behavior define a teacher's so-called style of teaching. It is reasonable to expect, then, that teacher behavior, as measured on the directness dimension, will reflect a teacher's style of teaching.

An association between measured teacher behavior and teacher style would be manifested in terms of a consistent distribution of tallies in the interaction matrix. Thus, for a particular teacher and group of students, it is reasonable to expect that beyond some relative minimum of observation, a consistency of tally distributions within the teacher's composite matrix will emerge. This implies that the matrix becomes more static with an increase in observation time. Hence, with additional observations, the proportions of tallies in the cells would undergo less and less change.

An attempt was made in this study to determine whether or not there is a connection between tally distribution and teaching style for high school physics teachers. Statistical techniques developed by Darwin were used to compare fall and spring matrices of individual teachers.

The Darwin Chi-Square Tests

In comparing two matrices, say A and B, the null hypothesis concerning the two distributions of tallies can be tested by the so-called Darwin Chi-Square Test. The statistic, a likelihood ratio criterion, has a sampling distribution of chi-square at $s(s-1)(r-1)$ degrees of freedom where s is the number of categories and r is the number of matrices compared. By calculating chi-square and referring to the sampling distribution at the appropriate degrees of freedom, one can determine the probability of the observed disparity in the distributions of tallies in A and B under the hypothesis that A and B are samples from the same distribution, i.e., that A and B are the same.

To calculate chi-square, a third matrix, say C, is formed by the cell-wise summation of A and B:

$$C_{ij} = A_{ij} + B_{ij}$$

for all rows, i, and columns, j. Then,

$$\text{chi-square} = 4.605(K - L - M + N)$$

where:

K = summation of n log n over each cell of A and B (n = tallies per cell, 200 terms)

L = summation of n log n over each row of A and B (n = tallies per row, 20 terms)

M = summation of n log n over each cell of C (n = tallies per cell, 100 terms)

N = summation of n log n over each row of C (n = tallies per row, 10 terms)

Logarithms are taken to the base 10. If the base e is used, the constant (4.605) is replaced with 2.

If more than two matrices are involved in a comparison, the same formulation is used involving each of the matrices and a summed matrix of all of them. In such a case, the test becomes one of homogeneity of matrices. All comparisons of interest in this study involved pairs of 10 X 10 matrices.

hence there were 90 degrees of freedom in each case.

For each teacher a comparison was made of fall and spring summed matrices corresponding to each of the Major Activities. For example, all of a teacher's observation data from the demonstration activity in the fall series of observations were put together to form a composite matrix. This matrix was compared to a similarly formed composite from the spring series of observations. If chi-square was less than or equal to 107.6, the matrices were considered to be not significantly different at the .10 level of confidence. Such a result would provide support to a contention that the distribution of tallies within the matrix is associated with behavioral consistencies, and hence teaching style of the teacher. The Wightman program was used with the Control Data 1604 computer for all calculations.

Table 18 shows the comparisons for Demonstration. The calculated value of chi-square and its significance are given. In addition, the number of tallies and observation time are reported. No test was possible for nine of the seventeen teachers; in these cases the activity was not observed in both fall and spring. For the eight tests possible, the null hypothesis could not be rejected at the .10 level in six cases. This suggests that the matrices reflect consistent aspects of teacher behavior in the Demonstration Activity.

Table 19 shows the results of tests carried out in the same fashion for Lecture. In this activity the matrices were not significantly different in ten out of fourteen cases. Teacher 06 showed marked differences, however.

Data from the Laboratory Activity was available in both series of observations on only three teachers. The comparisons are given in Table 20. The results are two out of three in favor of the matrix representing teacher style. But with only three cases, any conclusion would be questionable.

Table 21 shows comparisons for Recitation-Discussion. Out of fourteen comparisons, only seven pair were not significantly different. Of the seven cases in which the null hypothesis was rejected, the matrices were very different in six instances (probability less than .005).

Several points concerning the comparisons should be noted. First, these comparisons serve a secondary objective of the study. Hence, no effort was made to insure any minimum of observation time for each teacher in each activity. The data acquired in meeting Objective One (to describe what exists in terms of interaction analysis in the high school physics classroom) was used. Consequently, some comparisons involve matrices of many tallies being compared to those of relatively few tallies. In many cases no test was possible. From his studies of junior high school teachers, Flanders suggests a minimum of about three hours of observed interaction (about 3,600 tallies) to insure matrix stability. All comparisons here involved less than half that amount.

TABLE 18

DARWIN CHI SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY ONE - DEMONSTRATION

Teacher	Tot. Tallies Fall	Tot. Tallies Spg.	Time (min.) Fall	Time (min.) Spg.	Darwin Chi-Square[a]	Significance
01	0	66	0	4	------	no test
02	119	99	7	7	20.89	NS[b]
03	0	62	0	4	------	no test
04	0	1,164	0	59	------	no test
05	0	0	0	0	------	no test
06	202	0	11	0	------	no test
07	147	1,550	8	81	114.28	$.025 < p < .050$
08	638	0	32	0	------	no test
09	0	131	0	7	------	no test
10	816	610	41	30	90.50	NS
11	416	1,034	20	52	70.74	NS
12	0	752	0	37	------	no test
13	0	0	0	0	------	no test
14	300	686	15	35	43.93	NS
15	547	143	28	7	74.64	NS
16	73	731	4	37	48.28	NS
17	1,862	1,276	93	65	109.82	$.05 < p < .10$

[a] df = 90.

[b] Not significant at the .10 level, i.e., $p > 0.10$.

TABLE 19

DARWIN CHI SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY TWO - LECTURE

Tea-cher	Tot. Tallies		Time (min.)		Darwin Chi-Square[a]	Significance
	Fall	Spg.	Fall	Spg.		
01	1,341	0	67	0	------	no test
02	1,521	1,158	78	60	111.72	$.05 < p < .10$
03	1,085	1,662	55	85	122.43	$.010 < p < .025$
04	747	0	37	0	------	no test
05	752	663	37	33	93.27	NS[b]
06	1,128	1,526	56	76	129.21	$p < .005$
07	0	353	0	18	------	no test
08	973	950	48	48	101.37	NS
09	1,193	1,546	59	78	61.28	NS
10	1,144	1,650	57	83	77.57	NS
11	2,544	1,085	113	54	70.90	NS
12	756	1,144	38	57	95.13	NS
13	582	639	30	30	126.60	$.005 < p < .010$
14	958	1,709	48	85	98.91	NS
15	304	538	15	28	61.10	NS
16	801	819	41	41	91.34	NS
17	420	742	23	37	96.19	NS

[a] df = 90.

[b] Not significant at the .10 level, i.e., $p > 0.10$.

TABLE 20

DARWIN CHI SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY THREE - LABORATORY[a]

Teacher	Tot. Tallies		Time (min.)		Darwin Chi-Square[b]	Significance
	Fall	Spg.	Fall	Spg.		
02	1,127	1,906	60	96	119.09	$.010 < p < .025$
03	170	341	11	18	77.37	NS[c]
14	739	747	37	37	89.02	NS

[a] Comparison data available on three teachers only.

[b] df = 90.

[c] Not significant at the .10 level, i.e., $p > 0.10$.

TABLE 21

DARWIN CHI SQUARE COMPARISONS OF FALL AND SPRING MATRICES OF ACTIVITY FOUR - RECITATION

Teacher	Tot. Tallies		Time (min.)		Darwin Chi-Square[a]	Significance
	Fall	Spg.	Fall	Spg.		
01	1,090	1,539	56	78	160.84	$p < .005$
02	980	594	52	33	84.06	NS[b]
03	692	242	37	11	73.53	NS
04	1,619	903	81	44	237.68	$p < .005$
05	1,608	1,458	80	72	142.05	$p < .005$
06	528	249	26	13	110.48	$.05 < p < .10$
07	2,314	382	117	19	74.82	NS
08	1,411	2,304	70	115	143.40	$p < .005$
09	569	525	28	26	49.52	NS
10	48	0	3	0	------	no test
11	184	1,029	9	52	52.02	NS
12	1,502	213	76	11	94.12	NS
13	1,495	1,273	74	63	247.14	$p < .005$
14	237	0	11	0	------	no test
15	1,212	654	61	33	88.72	NS
16	1,331	651	67	33	136.03	$p < .005$
17	0	0	0	0	------	no test

[a] df = 90.

[b] Not significant at the .10 level, i.e., $p > 0.10$.

Second, comparisons of matrices involving very few tallies (e.g., Teacher 02, Table 18) are of questionable value. In such a case many of the cells have no tallies and exact agreement for a majority of cells results. This depresses the value of chi-square. Perhaps something like fifteen to twenty minutes should serve as a minimum if the comparison is to be at all useful.

In all comparisons reported here, a time lapse of about five months transpired between fall and spring observations. Thus, completely different physics topics are involved in the comparisons. Typically, teachers were in some phase of a unit on mechanics in the fall and dealt with electricity and magnetism during the spring observation.

Finally, it is noted that the Darwin test is very sensitive to fluctuations in the tally distribution. The intra-observer reliability tests showed a chi-square of 34 to be typical when the same data was recoded. Thus, some difference in distribution is due to coding error.

With consideration of these factors in conjunction with an examination of the Darwin test results, it seems reasonable to conclude that the distributions of tallies in these matrices are sufficiently static to be thought of as generally reflecting the physics teacher's style of teaching. The results support the contention most favorably in the Lecture Activity and least favorably in the Recitation Activity. Apparently physics teachers are more consistent in the way they develop new concepts through both lecture and demonstration than they are in the way they reinforce concepts already presented. In this regard no generalization concerning Laboratory can be presented.

An additional series of Darwin tests was carried out in which fall composites of all activities, excluding Laboratory, were compared to spring composites. These comparisons involve matrices on the order of 2,000 tallies in most cases. The results are given in Table 22. Interestingly enough the null hypothesis could not be rejected in only four out of seventeen cases. In twelve comparisons the matrices were very different. This result is in a direction opposite to what would be expected if only the effect of additional tallies is considered. The evidence seems to suggest that physics teachers are more consistent in their behavior within the Major Activities than they are overall. An implication is that physics teachers are likely to maintain a consistent style for a particular type of activity, but style is likely to be different from one activity to the next.

Teacher Influence and Student Attitude

In consideration of Objective One of the study it was felt that an assessment of student attitude would enhance the interaction analysis by providing an indication of how the student perceives the classroom situation. The Michigan

TABLE 22

DARWIN CHI SQUARE COMPARISONS OF SUMMED FALL MATRICES AND SUMMED SPRING MATRICES WITH ACTIVITY 3 (LABORATORY) EXCLUDED

Tea-cher	Tot. Tallies		Time (min.)		Darwin Chi Square[a]	Significance
	Fall	Spg.	Fall	Spg.		
01	2,431	1,605	123	82	248.03	$p < .005$
02	2,620	1,851	137	100	117.78	$.025 < p < .050$
03	1,777	1,966	92	100	160.59	$p < .005$
04	2,366	2,067	118	103	359.11	$p < .005$
05	2,360	2,121	117	105	166.06	$p < .005$
06	1,858	1,775	93	89	162.45	$p < .005$
07	2,461	2,285	125	118	247.77	$p < .005$
08	3,022	3,254	150	163	165.03	$p < .005$
09	1,762	2,202	87	111	68.31	NS[b]
10	2,008	2,260	101	113	104.37	NS
11	3,144	3,148	142	158	101.74	NS
12	2,258	2,109	114	105	105.98	NS
13	2,077	1,912	104	93	307.37	$p < .005$
14	1,495	2,395	74	120	155.93	$p < .005$
15	2,063	1,335	104	68	145.02	$p < .005$
16	2,205	2,201	112	111	140.66	$p < .005$
17	2,282	2,018	116	102	130.00	$p < .005$

[a] df = 90.

[b] Not significant at the .10 level, i.e., $p > 0.10$.

Student Questionnaire (MSQ) was modified[2] and entitled Student Questionnaire (SQ) for the purposes of this study.

The MSQ is an improved version of the Minnesota Student Attitude Inventory. The instrument consists of sixty four-alternative items.[3] The modifications of MSQ to produce SQ were limited to changes in the wording in cases where the language was considered inappropriate for twelfth graders. For example, "boys and girls" was replaced with "students." Student Questionnaire, a list oi the Michigan Student Questionnaire modifications made, and a memorandum giving instructions for the administration of SQ are given in Appendix C. SQ was considered valid only in the sense that its counterpart (MSQ) has been used in extensive research by Flanders at the elementary level.

It should be emphasized that the purpose of Student Questionnaire was to provide a rough indication of physics students' attitudes toward their class, the teacher, and school work in general. Particular interests were (1) to see if indirect influence is associated with more desirable student attitudes, and (2) to see how student perception of teacher directness, as measured by Student Questionnaire agrees with matrix results. Attitude scores were also used in tests of teacher effectiveness (later chapters). All of this is somewhat peripheral to the major objectives of the study. It was done as a matter of interest and for speculative purposes.

---

[2] The author is grateful to Dr. Ned A. Flanders of the University of Michigan for permission ot use MSQ as indicated above.

[3] Appended items, 61-64, are not part of SQ.

The Student Questionnaire was delivered by teachers and completed by students outside of class. Four class periods were reserved for other tests in this project; hence the consumption of an additional period for the administration of SQ was not desirable. The "Memorandum on Student Questionnaire" (Appendix C) was used to instruct teachers on procedures of administration. An addressed and stamped return envelope was provided for each student in the class observed. The instrument was given in early March 1965 so that an assessment of attitude near the middle of the school year when examination "pressure" is at a minimum would be possible. Through personal inquiries during the spring series of observations, the author was able to subjectively verify that all teachers followed the instructions explicitly.

Table 23 gives the results of Student Questionnaire returns. The per cent of the class responding extends from 50 to 100 for seventeen classes. Class mean SQ scores range from 139 to 195. It is noted that items are keyed in the direction of desirable attitudes, i.e., a higher score is associated with better attitudes. The mean of SQ class means is 171 and the median standard deviation lies between

TABLE 23

SCORES ON STUDENT QUESTIONNAIRE BY CLASSES[a]

Teacher	$\bar{X}$	SD	N	% of Class Responding	Rank
01	155	23	16	85.2	16.0
02	164	14	20	95.2	12.0
03	169	22	12	63.2	10.5
04	179	17	19	90.4	5.0
05	189	20	15	79.0	2.5
06	189	19	12	66.7	2.5
07	139	12	8	66.6	17.0
08	157	22	26	70.3	14.0
09	174	22	17	74.0	9.0
10	181	21	9	50.0	4.0
11	178	12	16	64.1	6.0
12	195	15	9	100.0	1.0
13	161	13	9	90.0	13.0
14	156	23	10	83.3	15.0
15	175	28	10	55.6	8.0
16	169	20	9	69.2	10.5
17	177	23	10	77.0	7.0

[a] Student Questionnaire is a slightly modified version of Michigan Student Questionnaire. The latter, an improved version of the Minnesota Student Attitude Inventory, was used by Flanders in studies of sixth-grade students. For 17 Regents Physics classes, the mean of SQ class means is 171 and the median standard deviation is between 20 and 21. Maximum possible score is 240 and minimum possible is 60.

20 and 21. These results compare favorably with Flanders' Michigan Student Questionnaire findings on 101 sixth-grade classes where the mean of class means was 178 with a median standard deviation between 21 and 22 (personal communication from Ned A. Flanders).

A Kendall tau correlation coefficient was calculated to determine correlations between class mean Student Questionnaire scores and various teacher and school characteristics. A correlation matrix involving numerous intercorrelations is presented in Chapter Eight. It suffices to mention here that class mean Student Questionnaire score correlated 0.19 with Total I/(I+D) for total teaching. However, when the Revised I/(I+D), based on total teaching was used, a tau of 0.43 resulted. This value is significantly greater than zero at the .05 level of confidence for a one-tailed test. The results support a contention that indirect influence used in motivation and control tends to be positively associated with better student attitudes. It also suggests little or no association between desirable attitudes and overall indirectness of teacher influence.

The same line of reasoning is possible concerning whether or not students perceive an indirect teacher to be indirect. In terms of the Revised I-D percentage, they do; for overall indirectness there is no perception one way or the other. These results are no more than suggested implications; they provide additional information relative to the matrix interpretations given earlier.

Two additional significant Kendall tau correlations were found. These were 0.38 with the Steady-State I/(I+D) for Total Teaching and 0.37 with the teacher's credit hours of preparation in science. Both are significant at the 0.05 level (one-tailed test).

Several additional points are appropriate. One could have more assurance in the results if the instrument was administered in the classroom. This was not practical, however. There is no way to be sure what difference in the results would have been found if all students responded. It is likely that some students may not have responded because of a hesitation to criticize their teacher. However, steps were undertaken to overcome this by means of complete anonymity. It seems just as likely that those who did not respond "forgot" or "just didn't get around to it." Nevertheless, the results are useful in the limited sense described.

Concluding Remarks

In this chapter the results representing the fulfillment of Objective One of the study are given. The store of normative teacher-student interaction data from high school physics classrooms was presented along with detail concerning various interpretations of the associated interaction matrices.

A grand matrix for physics teaching based on over seventy-two hours of observed interaction and containing over 86,000 tallies was presented and analyzed. Grand matrices for each of four Major Activities were also discussed. Aspects of the matrices were compared among activities and between physics teachers and junior high school teachers. In the latter case data from this study was compared to that reported by Flanders.

Various I-D percentage scores were examined. Tally loadings in individual matrix cells and patterns of loadings provided additional information. The entire effort led to a reliable description of what takes place in the high school physics classroom in terms of teacher-student interaction along a dimension of directness of teacher influence.

Some of the more important findings showed that physics teachers exhibit more direct influence but that much of this is due to an emphasis on content. However, less than 10 per cent of teacher talk is associated with concept development wherein social skills associated with motivational aspects of teacher behavior are employed. A particularly surprising result showed that teachers are more direct and less accepting of student ideas in the Laboratory Activity than in any other Major Activity. The implication is that teachers provide less opportunity for concept development through discovery in physics than might be imagined by those concerned with curriculum development.

Darwin chi-square comparisons of fall and spring composite matrices provided evidence to support a contention that the pattern of numbers in the matrix is associated with teacher behavioral consistencies. The implication is that the matrix reasonably represents aspects of a teacher's style of teaching.

Finally, measures of student attitude were discussed. The results provided additional information enhancing the development of a view of what takes place in the high school physics classroom.

The data presented will assist further research. In particular, it will facilitate implications concerning what changes may be desirable and how such changes could be effected through programs of teacher education.

CHAPTER SEVEN

DETERMINATION OF TEACHER EFFECTIVENESS

Introduction

A teacher is effective to the extent that he succeeds in producing desirable behavioral changes in his students as a consequence of teaching efforts. Student behavioral changes, assessed in this connection, are necessarily those which facilitate an inference concerning the extent to which the objectives of the teaching program have been met. It is clear, then, that effectiveness must be measured in terms of student achievement of program objectives; a teacher may be more effective with some objectives and less effective in terms of others. In order to determine objective attainment, measuring instruments must validly and reliably assess appropriate student behavior.

In this study three major objectives of high school physics were identified and an appropriate test to measure behavior congruent with objective attainment was selected in each case. Teacher effectiveness was assessed, for each objective, in terms of class mean achievement as measured. Linear multiple regression analyses were carried out to adjust student achievement scores for various aspects of initial ability and previous achievement.

Student Testing

Major Objectives Selected and Tests Used

With one exception (Teacher 08) the seventeen teachers and classes in the study were involved with the Regents Physics Course as currently outlined by the New York State Education Department. Teacher 08 taught a tentative version of the new Regents Course (Van Hooft, 1964) which will likely replace the old course throughout the State in 1967. It was felt that the major objectives to be identified and tested for should be of two types: (1) objectives that are basic to high school physics teaching in general, and (2) objectives that are basic to the New York State Regents Physics Course. With this approach teachers could be compared on bases of both meeting objectives associated with high school physics teaching in general and meeting objectives of the particular course syllabus used. Thus, conclusions reached might have extended generality.

Additional factors that influenced the identification of major objectives were (1) that the objectives are acceptable to teachers, administrators, science educators, and physicists as prime objectives of high school physics teaching; and (2) that relative attainment of the objectives by high school physics students is measurable so that teacher effectiveness in meeting the objectives could be determined.

Accordingly, three major objectives were established with associated criteria upon which to base teacher effectiveness:

1. Objective: Understanding of, and ability to use facts and principles of physics as specified in the minimum (required) syllabus of the New York State Regents Physics Course.

   Criterion of Teacher Effectiveness: Class mean adjusted score on New York State Regents Physics Examination, June 1965, New York State Department of Education.

2. Objective: Gain in understanding of, and ability to use facts and principles of physics commonly associated with high school physics courses in general.

   Criterion of Teacher Effectiveness: Class mean adjusted pre-post gain score on Cooperative Physics Test, Form Z (CPT), Cooperative Test Division, Educational Testing Service.

3. Objective: Gain in understanding of the nature of science in terms of extended understandings of science as an enterprise, the scientist as an individual, and the methods and aims of science.

   Criterion of Teacher Effectiveness: Class mean adjusted pre-post gain score on Test On Understanding Science, Form W (TOUS), by W. W. Cooley and L. E. Klopfer, Educational Testing Service, as follows:

Area I (TOUS-I)--Science as an enterprise.

Area II (TOUS-II)--The scientist as an individual.

Area III (TOUS-III)--Methods and aims of science.

TOUS-Total (I + II + III)--The nature of science.

Objectives 2 and 3 were considered as basic to high school physics teaching in general, and Objective 1 and 3 are considered fundamental to the Regents Physics Course.

It is noteworthy that although Objective 3 is an important objective of all areas of secondary science teaching, testing for student progress toward this goal is not common. Testing in the high school physics program is generally limited to attempts at measuring student understanding of subject matter facts and principles. Yet, one cannot doubt the equal importance of having students progress in their understandings concerning such things as the nature of science, what science is, what scientists do, and the role of science in our world today. It is reasonable (and necessary) to expect secondary science programs to contribute to these understandings. Very few students will become professional scientists--and a still smaller proportion will become physicists--but all are to be prepared to take their places in a free, democratic, and scientific society.

Validity and Reliability of Tests

A test is valid if in fact it measures what it purports to measure. The tests used in this study were considered to have satisfactory content validity, i.e., they represented in each case the consensus of educators concerning the kinds of things a student, who has met the objectives being measured, would know and would be able to do. In this sense the New York State Regents Physics Examination (RPE), the Cooperative Physics Test, Form Z (CPT), and Test On Understanding Science, Form W (TOUS) were accepted as valid instruments.

Secondary physics curricula are undergoing important changes throughout the country. Although Cooperative Physics Test may appear somewhat antiquated in light of present trends, it does adequately test for understandings of facts, principles, and concepts associated with the subject matter still commonly taught in high school physics courses. Devoid of items on "modern physics," it is a valid standardized test on "traditional physics." The latter is as yet commonplace--hence Cooperative Physics Test was chosen.

Cooley and Klopfer's report (1961) on their development of Test On Understanding Science at the Graduate School of Education, Harvard University, presents convincing evidence concerning the test's validity as a measuring instrument for Objective 3.

Finally, Regents Physics Examination, developed each year by the New York State Education Department, is considered to be a valid test measuring student understandings of facts, principles, and concepts of the New York State Regents Physics Course.

A test is reliable if there is evidence to support a conclusion that a score on the test is stable and trustworthy, i.e., that the score is a sufficiently good index of "true ability" in terms of student behavior that the test assesses. The reliability coefficient, r, in effect a correlation of the test with itself, serves as an index of the extent to which obtained score variance is free of "error variance, i.e., is a measure of "true score variance."

Data with which to estimate reliability on Regents Physics Examination was not available. Table 24 presents data on the reliabilities of Cooperative Physics Test and Test On Understanding Science available at the time of test selection.

Procedures of Test Administrations

Each participating teacher administered both Cooperative Physics Test and Test On Understanding Science pre-tests during the latter two weeks of October 1964. Post-testing of both was carried out during the middle two weeks of May 1965. Test booklets, answer sheets, and electrographic pencils were provided by the Principal Investigator (see Appendix E, Form E-1). Teachers were instructed to (1) administer Cooperative Physics Test first, (2) administer the tests in a

TABLE 24

CPT AND TOUS PUBLISHED RELIABILITIES

Test	Number of Items	r	$\bar{X}$	SD	N
Coop. Physics Test[a]	77	.82[b]	57.0	8.9	5000[c]
TOUS-I-Science as an enterprise	18	.58	----	----	2535
TOUS-II-The scientist	18	.52	----	----	2535
TOUS-III-Methods and aims of science	24	.58	----	----	2535
TOUS-Total-Nature of science	60	.76	----	3.49	2535

[a] Data reported for CPT pertains to "scaled scores"; see Burke, 1950.

[b] Calculated estimate of Kuder-Richardson r based on variance and mean of "scaled scores." Others are Kuder-Richardson.

[c] Grade 11 students from 90 high schools located in central and eastern United States.

secure fashion under normal testing conditions (as they would the Regents), (3) inform students that the tests were important but that test grades would not count toward their course grades, (4) urge students to do their best on the tests, (5) adhere strictly to the appropriate test time limits, and (6) not to provide any "make-ups" for absentees (see Appendix E, Form E-2). The same instructions were given to teachers for both fall and spring testing. A cover letter (Appendix E, Form E-3) reminded teachers that the test results would be provided in return.

No discrepancies in testing procedure were reported. It is reasonable to assume that all testing was properly conducted and that the method of test administration was constant for all classes. The Regents Physics Examination was administered as part of the regular school operation.

Reliability Determined From Test Results

Table 25 summarizes Cooperative Physics Test and Test On Understanding Science results. Reliabilities were calculated from pre and post scores to facilitate comparisons with published values. Comparisons were generally favorable with the exception of the TOUS-III pre-test scores.

The variation of "N" from pre to post testing was due to absenteeism, course drop-outs, rescheduling of some students at the close of the fall term, et cetera. The data is in terms of raw scores (as will always be the case unless

TABLE 25

CPT AND TOUS RELIABILITIES BASED ON PRE AND POST ADMINISTRATIONS

Test	No. Items (n)	Max. Possible Raw Score	Pre-test: Oct. 1964				Post-test: May, 1965			
			N	$\overline{X}$	SD	r^a	N	$\overline{X}$	SD	r^a
TOUS-I	18	18	321	11.9	2.9	0.55	275	11.5	2.8	0.50
TOUS-II	18	18	321	11.5	2.6	0.41	275	11.7	2.5	0.37
TOUS-III	24	24	321	12.4	2.7	0.19	275	12.9	3.0	0.35
TOUS-Total	60	60	321	34.7	6.8	0.70	294	35.8	7.0	0.72
CPT	77	77	328	26.1	8.9	0.83	301	37.6	10.1	0.88

[a] Approximation of Kuder-Richardson r; see formula, page 166.

otherwise noted). Class results of pre and post testing were reported to each teacher in early June (see Appendix E, Forms E-4 and E-5).

The reliability coefficient, r, is a sufficiently good approximation of the Kuder-Richardson reliability coefficient and is calculated by the relation given by Garrett, (1965, p. 341):

$$r = \frac{n(SD)^2 - \overline{X}(n-\overline{X})}{(SD)^2(n-1)}$$

where $\overline{X}$ is the mean score, SD is the standard deviation, and n is the number of test items. The above formula has the advantage of simplicity. The Kuder-Richardson r, which it approximates, in stressing intercorrelations of items within the test and correlations of items with the test score, is thought to be a "best" measure of test reliability. The formula could not be applied to Regents Physics Examination as its items are not equally weighted.

For the interpretation of reliability values it should be noted that the general consensus is that a correlation of ±0.40 to ±0.70 is considered to indicate a substantial or marked relationship. Correlations ranging from ±0.20 to ±0.40 are generally viewed as low--a relationship is present but slight.

It was concluded that the reliability of Cooperative Physics Test and Test On Understanding Science-Total was unquestionably sufficient for purposes of this study. Reliabilities of Test On Understanding Science part scores are substantially lower, especially Part III. Nevertheless, it was felt that there was sufficient reliability to warrant use of the part scores in estimations of teacher effectiveness. It is emphasized, however, that these lower reliabilities should be considered in reference to any conclusions of the study associated with Test On Understanding Science-I, II, and III. Although such conclusions must necessarily be given with less confidence, it was felt that implications connected with them might nevertheless be useful.

Student Achievement in Regents Physics

Teacher effectiveness was ultimately assessed in terms of class mean achievement associated with the objectives and evaluative criteria outlined earlier. Linear multiple regression analyses were carried out to adjust class means for student ability and previous achievement. In this section the statistical adjustment of test scores and resulting measures of student achievement will be described.

Initial Ability and Previous Achievement

The fall enrollment of the seventeen Regents Physics classes in this study consisted of 320 students. Table 26 presents aptitude and previous achievement characteristics of this group. These data were acquired directly from permanent records in the schools. It was anticipated that some of the measures would ultimately serve to statistically adjust Cooperative Physics Test and Test On Understanding Science pre-test scores and Regents Physics Examination scores to provide adjusted class mean scores. At the time of data collection there was no way of knowing which measures would be most useful for the regression analyses. The data are of additional interest in that they provide a picture of what the typical Regents Physics student is like academically. For this purpose the June 1965 Regents Physics Examination results were included.

Some interesting facts concerning Regents Physics students can be gleaned from the table. Of 320 initially enrolled, 296 took the final examination. The student typically goes into physics with Math 10 and biology in his background; 75 per cent had chemistry and Math 11 prior to physics. However, less than one-third had taken earth science. It is noted that these data were collected during the spring series of visits. The Math 12a course (43 per cent of the students) was taken concurrently with physics.

The interest of physics students in going to college is indicated by the percentage who took the New York State Regents Scholarship Examination (65 per cent).

Numerous IQ tests were used in the group of schools. Both reported IQ scores (raw) and stanines are given. The published mean and standard deviation of each test was used

TABLE 26

APTITUDE AND ACHIEVEMENT CHARACTERISTICS OF REGENTS PHYSICS STUDENTS[a]

Criterion	Mean Score	Standard Deviation	N	N/320
IQ Score (raw)	118	10.8	292	0.912
IQ-Stanine	7.44	1.25	292	0.912
CEEB-SAT - December 1964				
Verbal	524	93.8	175	0.547
Mathematics	592	82.7	175	0.547
NYS Regents Schol.[b]				
Aptitude	96.0	20.5	271	0.847
Achievement	92.9	16.7	271	0.847
Nat'l. Merit - 1964				
Math Usage	24.3	4.10	193	0.603
Natural Sci. Reading	23.8	3.78	193	0.603
Selection Score	114	17.7	206[c]	0.644
NYS Regents Examination[d]				
Math 10	80.6	10.7	299	0.934
Math 11	80.5	12.8	241	0.753
Math 12a	74.4	15.8	138	0.431
Earth Science	80.9	8.2	103	0.322
Biology	82.2	8.2	293	0.916
Chemistry	74.8	12.5	244	0.762
Physics - June 1965	74.9	12.9	296[e]	0.925
Mean of Biol. and Chem.	78.8	10.0	238	0.744

[a] Based on a sample of 320 students comprising the fall enrollment of 17 Regents Physics classes of 17 high schools in a 14-county central New York State region.

[b] New York State Regents Scholarship Examination, 1965.

[c] Part scores were not reported in one school.

[d] In cases of repeats, only the highest score was used.

[e] Thirty-two of 296 physics scores are scores on the Regents Physics examination for the Experimental Regents Physics Course.

to place the reported IQ's in terms of stanines.

One of the seventeen classes used the new Experimental Regents Physics Course. This class was deliberately included to enhance the ability of the sample to represent physics teachers and classes as they currently exist.

Finally, by looking at the means and standard deviations, one can get an idea of the score distributions. The necessary assumption of a normal distribution about the mean is reasonable in light of the tests and substantial N. For example, it is estimated that about 16 per cent of the Regents Physics Examination scores exceeded 88.

Pearson product-moment correlations were calculated to compare the June 1965 Regents Physics scores with each of the ability and previous achievement measures. Correlations with CPT and TOUS-Total pre and post scores and with TOUS-I, II, and III pre scores were done as well. TOUS part post scores were omitted. The purpose was to provide general information concerning associations of Regents Physics Examination with CPT and TOUS and to assist the identification of control variables, i.e., scores that would be useful in adjusting Regents Physics Examination for student ability and previous achievement. The results are organized in Table 27. It is noted that all calculations here, and elsewhere in the study where means, standard deviations, correlations, and regressions are involved, were done with the Control Data 1604 computer. In this case the CORMA program was employed (Searle, 1964a).

In one of the schools two physics classes were tested on *Cooperative Physics Test* and *Test On Understanding Science* (at the request of the teacher). Consequently, *CPT*, *TOUS*, and *Regents Physics* data were collected from eighteen classes. It is emphasized that the extra class is involved only in the data of Table 26. Most of the students in the eighteenth class were eleventh graders, hence little aptitude and achievement data was available from that group. It was decided to exclude that group from all results of the study except the correlations of Table 26. It was thought that the extra fifteen scores would enhance the correlation estimates in connection with *CPT* and *TOUS* scores.

The high correlations of *Regents Physics Examination* with *College Entrance Examination Board-Mathematics* (.62), with *New York State Regents Scholarship-Aptitude* (.64), and with the mean of *Biology* and *Chemistry Regents* (.62) are noteworthy. Particularly interesting is the .65 correlation of *Regents Physics Examination* with the *Earth Science Regents* score (although N is one-third as great here). This suggests *Earth Science Regents* score as the best predictor of *Regents Physics Examination*. However, it was not used as a control in view of the low N, i.e., only one-third of the *Regents Physics Examination* scores could have been adjusted in a regression analysis if it were used.

TABLE 27

PEARSON PRODUCT-MOMENT CORRELATIONS BETWEEN REGENTS PHYSICS (JUNE 1965) SCORES AND STUDENT APTITUDE AND ACHIEVEMENT SCORES[a]

Criterion	N	R
IQ Score (raw)	290	.374
IQ-Stanine	290	.358
CEEB-SAT - December 1964		
Verbal	168	.484
Mathematics	168	.620
NYS Regents Schol.[b]		
Aptitude	250	.639
Achievement	250	.593
Nat'l. Merit - 1964		
Math Usage	181	.511
Natural Sci. Reading	181	.510
Selection Score	194[c]	.542
NYS Regents Examination[d]		
Math 10	296	.458
Math 11	229	.453
Math 12a	133	.508
Earth Science	93	.652
Biology	288	.498
Chemistry	231	.597
Mean of Biology and Chemistry	225	.624

(continued)

TABLE 27 (continued)

Criterion	N	R
Coop. Physics Test		
Pre-test: October 1964	301	.445
Post-test: May 1965	280	.641
TOUS Total Score		
Pre-test: October 1964	295	.398
Post-test: May 1965	276	.421
TOUS Parts (Pre-test)		
I. Science as an enterprise	295	.301
II. The scientist	295	.264
III. Methods and aims of science	295	.415

[a] Based on 311 students who took the Physics Regents. These students belong to a group of 338 students comprising the fall enrollment of 18 Regents Physics classes of 17 high schools in a 14-county central New York State region. Thirty-two of the 311 Regents scores are scores on the Regents Physics Examination for the Experimental Regents Physics Course.

[b] New York State Regents Scholarship Examination, 1965.

[c] Part scores were not reported in one school.

[d] In cases of repeats, only the highest score was used.

It is interesting to note that of all the aptitude and previous achievement variables, both forms of the IQ score are individually least correlated with Regents Physics Examination. Frequently, IQ is just "blindly" used as a control for achievement comparisons. These results suggest a fallacy in such practices and point up the necessity of exploring a wide range of control variables in this study. (One wonders how often students are led away from physics in view of an IQ score.)

Cooperative Physics Test and Test On Understanding Science pre scores were included in these correlations as they represent aspects of previous achievement. The two post measures were included as a matter of interest. The CPT and TOUS post results show that to some extent, the three tests measured the same kinds of things. In the case of CPT and Regents Physics Examination the extent was greater. Although a hindrance, no impairment of the usefulness of the tests for purposes of this study was implied.

Selection of Control Variables

Data on fifteen selected control variables were available for an investigation concerning which combination of variables would best serve the adjustment of achievement criteria for student initial ability and previous achievement. In the cases of Cooperative Physics Test and Test On Understanding Science pre-test scores were to be adjusted to provide adjusted pre-post gains. The problem was complicated by the fact that not all students had scores on all the control variables. A complete set of data on each of the fifteen control variables was available for only 81 of the 320 initially enrolled students.

It is clear that any statistical method of score adjustment involves error. Score adjustments through regression analyses are generally improved with increased numbers of control variables and increased numbers of scores (N). In this study any attempt to include more control variables necessitated a decrease of N. Thus, it was necessary to undertake analyses to determine which combination of the fifteen available variables would provide the best adjustment for all evaluative criteria (i.e., _Regents Physics Examination_, _Cooperative Physics Test_, and _Test On Understanding Science_). One set of controls for all adjustments was more practical than a variation of methods among the criteria.

A Pearson product-moment correlational analysis was carried out using 81 sets of scores on the fifteen control variables. The inter-correlation matrix is given in Table 28. The variables as listed and defined in the legend are designated as X_1, X_2, X_3, et cetera, in subsequent tables, i.e., _New York State Regents Scholarship-Aptitude_ is X_1. The correlation matrix was used to identify variables that were highly correlated with _TOUS_-Total pre-test and not

TABLE 28

INTER-CORRELATION COEFFICIENTS OF 15 CONTROL VARIABLES AND TOUS-TOTAL PRE-TEST BASED ON 81 SETS OF OBSERVATIONS[a]

	Apt	Ach	RA	RCE	RBE	RB	IQr	IQs	CBv	CBm	NMm	NMsr	NMss	M10	M11	TOUS-T
Apt	1	819	454	524	623	632	552	523	824	711	586	583	786	448	454	594
Ach		1	469	540	664	668	601	542	789	607	496	509	739	442	431	643
RA			1	633	499	660	421	384	427	549	509	343	480	696	658	451
RCE				1	475	908	277	272	467	509	458	354	500	397	428	324
RBE					1	775	356	327	660	393	330	534	615	385	275	470
RB						1	350	334	615	519	452	490	626	437	401	442
IQr							1	892	565	537	589	420	649	513	448	537
IQs								1	503	485	583	343	623	473	476	456
CBv									1	558	441	607	796	380	393	669
CBm										1	691	470	611	582	527	456
NMm											1	365	658	635	619	479
NMsr												1	786	313	209	613
NMss													1	528	469	732
M10														1	668	445
M11															1	412
TOUS-T																1

(See overleaf for legend)

TABLE 28 (continued)

Legend

Symbol	Definition	Symbol	Definition
Apt	New York State Regents Scholarship Examination--Aptitude; 1965 series	CBv	CEEB-SAT, Verbal, December 1964
Ach	New York State Regents Scholarship Examination--Achievement; 1965 series	CBm	CEEB-SAT, Mathematics, December 1964
RA	Regents Index (Accounts for courses taken as well as grades)	NMm	National Merit Scholarship, Math Usage, Fall 1964
RCE	Regents Chemistry Examination	NMsr	National Merit Scholarship, Natural Science Reading, Fall '64
RBE	Regents Biology Examination	NMss	National Merit Scholarship, Selection Score, Fall 1964
RB	Mean of Regents Chemistry and Regents Biology	M10	Regents Mathematics Ten
IQr	IQ raw score (Score reported on permanent records; various tests and years involved)	M11	Regents Mathematics Eleven
IQs	IQr converted to stanines	TOUS-T	Test On Understanding Science, Total Score, pre-test, October 1964

[a] Pearson product-moment correlations based on 81, out of 320 participants, who had scores on each of the 16 variables. The students represent 17 Regents Physics classes in 17 high schools of central New York State.

highly correlated with each other. Such variables would be more useful as adjusters (controls).

Test On Understanding Science-Total pre-test was used to represent the effectiveness criteria. It was felt that the best controls for TOUS-Total would likely provide the best adjustment of Cooperative Physics Test pre-test scores and of Regents Physics Examination. Although adequately reliable, TOUS-Total was a little less reliable than CPT. Also, TOUS-Total unadjusted gains were relatively small compared to CPT gains. Finally, CPT and Regents Physics Examination are moderately correlated. These facts provided cause for a decision to develop a scheme of adjusting TOUS-Total and then to use that method for CPT and Regents Physics Examination as well.

The matrix must be examined in terms of N as well as r. Reference to Table 26 will assist comparisons. For example, College Entrance Examination Board-Verbal scores have a high correlation (.669) with Test On Understanding Science, but only 55 per cent of the students have CEEB scores. This low N weighs against using CEEB as a control. The IQ raw scores correlate .537 with TOUS and such data is available for 91 per cent of the students. However, reference to Table 27 shows that this variable would not be very useful in adjusting Regents Physics Examination scores. By carefully examining the data of the three tables, with r's and N's considered for both TOUS and Regents Physics Examination, a decision was reached to consider only six variables as possible controls. These are the first six variables listed in Table 28. (The given arrangement was designed after the selection was completed.)

It remains to point out that variable "RA," a so-called "Regents Index," is a composite score specially designed for purposes of this study. Its design was a purely speculative venture based on a "hunch" that it might be useful. It was initially endowed with a distinct advantage that all students would have an RA score. The following outlines the scheme for calculating a student's RA score:

1. For Math 10, 11, and 12a; and Biology; and Chemistry Regents scores; the following points are assigned:

    4 if in the 90's

    3 for 75-89 inclusive

    2 for 60-74 inclusive

    1 for 50-59 inclusive

    0 for less than 50 or no score

2. If the student is repeating physics, he is credited with it as in Item 1 above. For repeats in courses other than physics, only highest score earned was used.
3. If a student reports courses from out of state, he is credited with them as in Item 1 above.
4. Intermediate Algebra and Advanced Algebra Regents scores each get one-half the value on the scale of Item 1 above. This holds only when both are half-year courses.
5. If the student had Regents Earth Science, assign 2 points if the Regents score is 75 or higher; 1 point if 50-74, inclusive; and 0 if less than 50.

6. The one-year Intermediate Algebra Course is credited as in Item 5 above.
7. If Math 12b and no Math 12a, credit as in Item 1 above.
8. If both Math 12a and Math 12b, then Math 12a is credited normally (Item 1) and Math 12b is treated according to Item 5 above.

The detail of the above scheme was necessary to account for all physics students in the study. This points up the variety of backgrounds involved. It should be noted that in accounting for both grades earned and courses taken, RA is sensitive to both quantity and quality of preparation. Its correlation with TOUS-Total (.451) is better than Regents Chemistry (.324) and as good as Math 10 (.445) and Math 11 (.412). It was more useful than the latter two in view of N.

Regression Analyses

A series of linear multiple regression analyses was carried out using the MUREG Program (Searle, 1964b) on the 1604 computer. It should be clear that in the cases of TOUS and CPT, linear multiple regressions of the pre-test scores were made on selected control variables. The calculated regression lines were then used to produce an adjusted pre-test score which takes initial ability and previous achievement into account. The adjusted scores were subtracted from post-test scores to get adjusted gain scores. In the case of the Regents Physics Examination no gain is involved. Regents

Physics Examination is considered an achievement level score; it was adjusted for ability and previous achievement in the same way that pre-test scores were adjusted.

For all regression analyses, the multiple correlation coefficient, R and R^2 were calculated. In addition, $S_{Y_{est}}$, the standard error of estimated values of the dependent variable, Y, are given. The estimated values (Y_{est}) are determined by the regression equation. It is noted that:

$$S_{Y_{est}} = S_Y \sqrt{1 - R^2}$$

where S_Y is the standard deviation of known Y values. Also, R is the best possible correlation from a linear equation relating actual and predicted scores. It indicates how accurately scores derived from the regression equation represent actual values of Y.

With perfect linear multiple correlation, R^2 equals one and $S_{Y_{est}}$ equals zero. In such a case the regression equation would be a perfect predictor, i.e., for each value of each independent variable, there would be perfect agreement between the actual Y and the estimated Y. On the other hand if R^2 equals zero, then $S_{Y_{est}}$ equals S_Y and the regression equation is said to account for none of the variance of Y. In other words, a "blind" guess would provide as good an estimate of Y for particular values of the X's as would the regression equation.

The program provided for both direct and step-wise regression. In direct regression all independent variables in the input are used. In step-wise regression the independent variable which is most effective in accounting for the variation in Y, say X_1, is used first. Then the regression is repeated using the next most useful variable, and so on. In each case an F test is used to determine whether or not the subsequently calculated regression equation provides a significantly higher precision in predicting Y values. Thus, the process works through the X_i yielding a higher R at each repetition until R no longer increases. At that point remaining X_i are discarded. The method is often referred to as the "Wherry-Doolittle Test," (see Garrett, 1965, p. 426).

The F test involved is an analysis of variance for regressing the dependent variable, Y, on k independent variables with N observations each. The F ratio is MSR/MSE at k and n-k-1 degrees of freedom where MSR is the mean square about regressions, i.e., SSR/k where SSR is the regression sum of squares. MSE is the error mean square, i.e.:

$$\mathrm{MSE} = \mathrm{SSE}/(n-k-1)$$

where SSE, the error sum of squares is given by the relation

$$\mathrm{SSE} = \sum (Y_{est} - Y)^2$$

Again, note that Y_{est} values are predicted values derived from the regression equation while Y values are measured values used in conjunction with the corresponding values of the X_i's to derive the regression equation. Note also, that:

$$SST = SSE + SSR$$

Table 29 gives the results of three linear multiple regressions of TOUS-Total pre-test scores on various combinations of control variables. In each case the number of observations was 81. Note that the table is set up to show the various sums of squares, mean sums of squares, degrees of freedom, and F. Multiple R, R^2, and S_{Yest} are given as well. In addition the 99.5 percentile value of F is given to provide an indication at a glance of how F compares to the critical value for a one-tailed test at the .005 level of confidence. In effect, the test shows whether or not R is significantly different from zero.

In comparing the various cases one should also examine the S_{Yest} values. Again, the desirable direction is a maximum R and a minimum S_{Yest}. The X_i ($i = 1, 2, 3$, etc.) may be identified by reference to Table 28 (X_1 is the first variable listed, X_2 is the second, and so on).

As can be seen from Table 29, Case 1 in which all fifteen variables are used is best. Here S_{Yest} is 3.60 while S_Y is 5.45. Note that R equals .797 is a maximum for this case. The regression was step-wise. One can determine from the table how many variables were ultimately used by examining the df values. Note that in Case 1, k equals 13, the number of variables used. Thus, in this case, after

TABLE 29

LINEAR MULTIPLE REGRESSION ANALYSES OF TOUS-TOTAL PRE-TEST ON VARIOUS COMBINATIONS OF CONTROL VARIABLES USING 81 OBSERVATIONS[a]

	Sum of Squares (SS)	df	Mean (SS/df)	F^b	$F_{.995}$	R	R^2	S_{Yest}
Case 1: stepwise regression using 15 controls.								
Total	2378	80	----					
Regr.	1512	13	116.3	8.99	2.57[c]	.797	.636	3.60
Error	866	67	12.9					
Case 2: stepwise regression using X_1, X_2, X_3, X_4, X_5 and X_6.								
Total	2378	80	----					
Regr.	1137	6	189.5	11.3	3.49[d]	.691	.478	4.01
Error	1241	74	16.8					
Case 3: direct regression using X_1, X_2, and X_3.								
Total	2378	80	----					
Regr.	1072	3	357.5	21.1	4.63[e]	.672	.451	4.12
Error	1306	77	17.0					

---

[a] For this group, TOUS-Total pre-test has a mean of 37.6 and a standard deviation of 5.45.

[b] In each case, significant beyond the .005 level.

[c] df = 15 and 60.

[d] df = 6 and 60.

[e] df = 3 and 77; calculated by interpolation.

thirteen variables were used, the remaining two variables provided no additional information that was useful.

In Case 2 with six variables and 81 observations, F is greater but so is $F_{.995}$. More importantly, S_{Yest} is greater (on the order of 10 per cent) and R has decreased. The difference was substantial. When only the first three variables are used (Case 3), as opposed to the first six, the difference is slight.

The reader is reminded that X_1, X_2, and X_3 ultimately formed the "best" combination and were used to produce the adjusted scores. This is why, for this report, they were given the subscripts 1, 2, and 3.

Table 30 shows the results using 203 observations on the first six variables and then on the first three. By disregarding nine of the fifteen variables, N was increased from 81 to 203. Note in Case 1 that the step-wise regression selected only four of the six variables $(k = 4)$. This indicated that two of the six were unnecessary. There is little difference between using six and three as in the previous set of cases. Although Case 1 of Table 30 appears to be a considerably poorer regression than Case 2 of Table 29, S_Y for 203 observations is greater than S_Y when N equals 81. The ratios of S_{Yest} to S_Y are not very different (.736 at $N = 81$, and .766 at $N = 203$). A higher ratio implies a poorer "fit" since the difference between S_Y and S_{Yest} is associated with the amount of variance in Y accounted for by the regression equation.

TABLE 30

LINEAR MULTIPLE REGRESSION ANALYSES OF TOUS-TOTAL PRE-TEST USING 203 OBSERVATIONS[a]

	Sum of Squares (SS)	df	Mean (SS/df)	F[b]	$F_{.995}$	R	R^2	S_{Yest}
Case 1: stepwise regression using X_1, X_2, X_3, X_4, X_5 and X_6.								
Total	7725	202	----					
Regr.	3290	4	822.5	36.7	3.92[c]	.653	.426	4.73
Error	4435	198	22.4					
Case 2: direct regression using X_1, X_2, and X_3.								
Total	7725	202	----					
Regr.	3229	3	1076	47.8	4.50[d]	.646	.418	4.75
Error	4496	199	22.6					

[a] For this group, TOUS-Total pre-test has a mean of 35.7 and a standard deviation of 6.18.

[b] In each case, significant beyond the .005 level.

[c] df = 4 and 120.

[d] df = 3 and 120.

The detail of the foregoing is given to clarify both the approach that was taken in determining the best way to adjust the test scores and to guide the reader in extracting information from the tables.

Tables 31 and 32 respectively give corresponding analyses concerning Cooperative Physics Test pre-test and Regents Physics Examination scores. Note that in these instances if the first six variables were to be ultimately used, only 208 CPT and 178 RPE scores could be adjusted.

The results of these analyses made it clear that the best alternative was to use three independent variables, the New York State Regents Scholarship Examination-Aptitude, the NYSRSE-Achievement, and the Regents Index score, RA. This necessitated omitting eleventh grade students from the study. As a consequence teacher effectiveness scores are in terms of meeting the objectives (outlined earlier) with twelfth grade students. In thirteen of the seventeen classes, all students were twelfth graders. In one case 20 per cent were in grade 11 (Teacher 09's class, 5/25). In the remaining three classes the proportions of eleventh graders were as follows: Teacher 06, 7/19; Teacher 07, 11/16; and Teacher 17, 7/13. It appeared that no more than two classes would be lost in the teacher effectiveness phase of the study. This would leave fifteen classes. The original design called for fifteen classes with an additional two as a "margin."

Linear multiple regression analyses were run using

TABLE 31

LINEAR MULTIPLE REGRESSION ANALYSES OF CPT PRE-TEST USING 208 OBSERVATIONS[a]

	Sum of Squares (SS)	df	Mean (SS/df)	F^b	$F_{.995}$	R	R^2	S_{Yest}
Case 1: stepwise regression using X_1, X_2, X_3, X_4, X_5 and X_6.								
Total	16377	207	----					
Regr.	4027	4	1007	16.6	3.92[c]	.496	.246	7.80
Error	12350	203	60.8					
Case 2: direct regression using X_1, X_2, and X_3.								
Total	16377	207	----					
Regr.	3554	3	1185	19.8	4.50[d]	.466	.217	7.93
Error	12823	204	62.9					

[a] For this group, CPT pre-test has a mean of 27.0 and a standard deviation of 8.90.

[b] In each case, significant beyond the .005 level.

[c] df = 4 and 120.

[d] df = 3 and 120.

TABLE 32

LINEAR MULTIPLE REGRESSION ANALYSES OF PHYSICS REGENTS (JUNE 1965) USING 178 OBSERVATIONS[a]

	Sum of Squares (SS)	df	Mean (SS/df)	F^b	$F_{.995}$	R	R^2	S_{Yest}
Case 1: stepwise regression using X_1, X_2, X_3, X_4, X_5 and X_6.								
Total	25482	177	----					
Regr.	12845	4	3211	44.0	3.92[c]	.710	.504	8.55
Error	12638	173	73.1					
Case 2: direct regression using X_1, X_2, and X_3.								
Total	25482	177	----					
Regr.	12146	3	4049	52.9	4.50[d]	.690	.477	8.75
Error	13337	174	76.6					

[a] For this group, Physics Regents (June 1965) has a mean of 75.6 and a standard deviation of 12.0.

[b] In each case, significant beyond the .005 level.

[c] df = 4 and 120.

[d] df = 3 and 120.

all twelfth grade students having scores on X_1, X_2, X_3 and the various dependent criteria. Table 33 summarizes the results. The table gives the same kind of information as the other regression tables; however, the multiple regression equations are included. In addition, the TOUS part scores (I, II, and III) were run in this final series.

In each case Y' is the predicted value of the criterion. The sub-numbers distinguish the criteria. Note that of the three TOUS parts, II has the lowest multiple R. Note also that the adjustment of TOUS-Total was better than the adjustment of any of the TOUS part scores. A contributing factor is the fact that TOUS-Total score is more reliable than any of the part scores. However, although Cooperative Physics Test had a higher reliability than Test On Understanding Science-Total, the latter yielded a greater multiple R. Nevertheless, since the highest multiple R was in the case of Regents Physics Examination (Case 6), it seems reasonable that RPE was at least as reliable as the others. (Recall that r could not be calculated for RPE.)

Teacher Effectiveness

The adjusted Regents Physics Examination class means are listed in Table 34. Note that Teachers 07 and 17 have been deleted because the majority of the class did not have adjusted scores (eleventh grade students and/or no X_1 and X_2). Teacher 08's class used the new experimental syllabus. The

TABLE 33

LINEAR MULTIPLE REGRESSION ANALYSES USED TO ADJUST TOUS AND CPT PRE-TEST AND PHYSICS REGENTS (JUNE 1965) SCORES WITH NYSRS-APT. (X_1), NYSRS-ACH. (X_2) AND REGENTS INDEX (X_3) AS INDEPENDENT VARIABLES

Sum of	Squares (SS)	df	Mean (SS/df)	F[a]	$F_{.995}$	R	R^2	$S_{Y_{est}}$
Case 1: Adjustment of TOUS-I pre score (Y_1); 251 observations with a mean of 10.9 and a standard deviation of 2.92.								
Total	2130	250	----					
Regr.	779	3	260	47.6	4.50[b]	.605	.366	2.34
Error	1350	247	5.47					
(Equation: $Y_1' = 1.24 + 0.0195X_1 + 0.0836X_2 + 0.00852X_3$)								
Case 2: Adjustment of TOUS-II pre score (Y_2); 251 observations with a mean of 11.4 and a standard deviation of 2.62.								
Total	1717	250	----					
Regr.	412	3	137	25.9	4.50[b]	.490	.240	2.30
Error	1305	247	5.28					
(Equation: $Y_2' = 4.84 + 0.0186X_1 + 0.0428X_2 + 0.0685X_3$)								
Case 3: Adjustment of TOUS-III pre score (Y_3); 251 observations with a mean of 12.5 and a standard deviation of 2.82.								
Total	1987	250	----					
Regr.	558	3	186	3.22	4.50[b]	.530	.281	2.40
Error	1429	247	5.78					
(Equation: $Y_3' = 4.34 + 0.00753X_1 + 0.0769X_2 + 0.0254X_3$)								

(continued)

TABLE 33 (continued)

	Sum of Squares (SS)	df	Mean (SS/df)	F^a	$F_{.995}$	R	R^2	S_{Yest}
Case 4: Adjustment of TOUS-Total pre score (Y_4); 251 observations with a mean of 34.8 and a standard deviation of 6.65.								
Total	11058	250	----					
Regr.	5108	3	1703	70.6	4.50^b	.680	.462	4.91
Error	5950	247	24.1					
(Equation: $Y_4' = 10.4 + 0.0460X_1 + 0.203X_2 + 0.102X_3$)								
Case 5: Adjustment of CPT pre score (Y_5); 256 observations with a mean of 26.0 and a standard deviation of 9.13.								
Total	21246	255	----					
Regr.	5525	3	1842	29.5	4.50^b	.510	.260	7.90
Error	15721	252	62.4					
(Equation: $Y_5' = 0.446 + 0.122X_1 + 0.153X_2 - 0.0304X_3$)								
Case 6: Adjustment of Physics Regents (Y_6); 216 observations with a mean of 73.7 and a standard deviation of 12.9.								
Total	35616	215	----					
Regr.	19008	3	6336	80.9	4.50^b	.731	.534	8.85
Error	16607	212	78.3					
(Equation: $Y_6' = 33.7 + 0.215X_1 + 0.0458X_2 + 1.30X_3$)								

[a] In each case, significant beyond the .005 level.

[b] df = 3 and 120.

TABLE 34

CLASS MEAN ADJUSTED SCORES ON PHYSICS REGENTS (JUNE 1965)

Teacher	$\overline{X}$	SD	N	95% Confidence Limits
01	78.4	7.8	18	78.4 ± 3.9
02	73.9	10.1	21	73.9 ± 4.6
03	71.6	7.7	20	71.6 ± 3.6
04	72.6	5.2	19	72.6 ± 2.5
05	72.4	9.5	18	72.4 ± 4.7
06	71.7	9.2	11	71.7 ± 6.2
09	67.8	11.9	13	67.8 ± 7.2
10	75.9	11.1	16	75.9 ± 5.9
11	76.2	11.5	23	76.2 ± 5.0
12	65.7	8.6	9	65.7 ± 6.6
13	78.8	8.8	6	78.8 ± 8.2
14	78.8	10.3	9	78.8 ± 8.0
15	74.6	6.2	17	74.6 ± 3.2
16	75.6	8.8	9	75.6 ± 6.8

standard deviations and 95 per cent confidence limits are given. The range of class means extends from 65.7 to 78.8.

Tables 35, 36, and 37 present class means of adjusted pre-post gain scores on Cooperative Physics Test, Test On Understanding Science-Total, and TOUS parts respectively. Gains were determined by subtracting adjusted pre-test scores from actual post-test scores to get adjusted gains. Class mean adjusted gains on CPT range from 4.2 to 16.9 raw score points. Ninety-five per cent confidence limits are given only for CPT and TOUS-Total.

An examination of the TOUS-Total adjusted gains shows remarkably little gain. Note that in six out of fifteen classes a loss is reported. The "gains" extend from -4.9 to 3.0. Maximum possible score on this test is 60.

For the group of all the students in the study the mean TOUS-Total pre-test score was 34.7 and the mean post-test score was 35.8. The range of gains is such that attempts to distinguish more effective teachers from less effective teachers on the basis of this criterion were still feasible. But the fact that overall, students typically made no gain on this test, is quite disturbing. The test is reliable and valid; the objectives it tests for are of prime importance in science education. The view that these objectives are met right along with subject matter objectives is placed in serious question. The same conclusion concerning TOUS part scores is unavoidable, even though the latter were less reliable measures than TOUS-Total.

TABLE 35

CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON COOPERATIVE PHYSICS TEST, FORM Z (CPT)[a]

Teacher	$\bar{X}$	SD	N	95% Confidence Limits
01	11.1	9.9	16	11.1 ± 4.3
02	9.8	8.5	20	9.8 ± 4.0
03	10.1	7.4	14	10.1 ± 4.2
04	16.1	9.4	18	16.1 ± 4.7
05	11.1	8.8	17	11.1 ± 4.6
06	4.2	8.4	9	4.2 ± 6.5
08[b]	9.8	8.7	25	9.8 ± 3.6
09	13.5	9.0	11	13.5 ± 6.0
10	16.9	8.9	16	16.9 ± 4.7
11	9.6	10.2	16	9.6 ± 5.4
12	14.3	10.8	8	14.3 ± 9.0
13	9.0	7.0	8	9.0 ± 5.8
14	9.8	9.2	8	9.8 ± 7.7
15	14.5	6.9	14	14.5 ± 4.0
16	11.2	7.3	8	11.2 ± 6.1
17	9.4	8.0	5	9.4 ± 9.9

[a] Raw scores.

[b] Physics course based on the new experimental State syllabus.

TABLE 36

CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TEST ON UNDERSTANDING SCIENCE, FORM W, (TOUS) TOTAL SCORE

Teacher	$\bar{X}$	SD	N	95% Confidence Limits
01	2.8	3.0	16	2.8 ± 1.6
02	-0.9	4.5	17	-0.9 ± 2.3
03	-2.3	6.5	11	-2.3 ± 4.4
04	-0.3	7.0	17	-0.3 ± 3.6
05	-0.3	4.3	16	-0.3 ± 2.3
06	1.1	5.3	9	1.1 ± 4.1
08[a]	1.5	5.0	22	1.5 ± 2.2
09	2.6	2.3	8	2.6 ± 1.9
10	2.9	3.9	15	2.9 ± 2.2
11	2.9	5.5	17	2.9 ± 2.8
12	0.2	3.6	8	0.2 ± 3.0
13	3.0	5.5	7	3.0 ± 5.1
14	-0.9	4.6	8	-0.9 ± 3.8
15	2.7	3.5	15	2.7 ± 1.9
16	-4.9	7.8	9	-4.9 ± 6.0

[a] Physics course based on the new experimental State syllabus.

TABLE 37

CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TEST ON UNDERSTANDING SCIENCE, FORM W, (TOUS) PARTS I, II AND III[a]

Teacher	N	TOUS-I		TOUS-II		TOUS-III	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
01	16	1.0[c]	2.3	0.2	1.6	0.7	2.3
02	17	0.0	2.1	-0.2	1.8	-0.7	2.8
03	11	-0.7	2.2	-1.1	2.5	-0.7	4.0
04	17	-0.2	2.8	-0.1	2.7	-0.8	2.9
05	16	-0.1	2.6	-0.7	3.0	-0.4	2.1
06	9	-0.7	2.9	1.1[c]	1.8	0.3	1.7
08[b]	22	0.4	2.3	0.0	2.0	0.8	3.0
09	8	1.5[c]	1.8	0.8	2.0	0.5	2.0
10	15	1.3[d]	1.9	0.7	2.1	1.0	2.8
11	17	0.3	2.4	0.7	3.0	1.7[d]	2.2
12	8	0.4	1.2	-0.6	2.9	-0.2	2.0
13	7	1.9[c]	2.5	0.1	2.1	1.0	2.0
14	8	0.1	2.5	-0.9	1.6	-0.1	2.7
15	15	1.7[d]	1.9	0.7	1.6	0.3	1.8
16	9	-1.6	3.1	-1.2	3.3	-2.0[c]	3.1

[a] I-Science as an enterprise; II-The scientist as an individual; III-Methods and aims of science.

[b] Physics course based on the new experimental State syllabus.

[c] Significantly different from zero at the .05 level.

[d] Significantly different from zero at the .01 level.

The grand summed matrix of all teachers in all activities (Table 10, Chapter Six) was produced to provide a Flanders matrix of the typical Regents Physics teacher. The data of the matrix is considered normative in view of the experimental design. By combining all students for whom adjusted achievement scores on Regents Physics Examination, Cooperative Physics Test, and Test On Understanding Science were available, an approximate measure of the effectiveness of the typical teacher results. This effectiveness is in terms of the course objectives measured. The results are "somewhat normative" in view of unequal numbers of students representing each teacher. However, all students in this group were in grade 12, and were statistically controlled for initial ability and previous achievement.

Table 38 gives comparisons of grand means of adjusted student achievement scores, or what is the same, teacher effectiveness scores. Inter-correlations of these scores are provided in Table 39. The Pearson r's support the contention that the course objectives measured are different, i.e., that the teacher effectiveness scores are different. Clearly, TOUS parts measure much of the same kind of things as TOUS-Total. Hence, the teacher effectiveness scores compared in Table 38 are of three different types. That is, in this study, teachers are in fact effective in three different ways as anticipated when the experimental design was constructed.

TABLE 38

GRAND MEANS OF ADJUSTED PHYSICS REGENTS SCORES AND OF ADJUSTED CPT, TOUS-I, II, III, AND TOUS-TOTAL PRE-POST GAINS

Test	$\bar{X}$	SD	N	95% Confidence Limits
Physics Regents (June 1965)	73.8	9.4	216[a]	73.8 ± 1.3
Coop. Physics Test (raw scores)	11.1	10.4	216[b]	11.1 ± 1.4
TOUS-I-Science as an enterprise	0.4	2.4	202[b]	0.4 ± 0.3
TOUS-II-The scientist	0.1	2.3	202	0.1 ± 0.3
TOUS-III-Methods and aims of science	0.1	2.6	202	0.1 ± 0.4
TOUS-Total-Nature of science	0.8	5.2	202	0.8 ± 0.7

[a] Grade 12 Regents Physics students of 16 teachers in 16 central New York State high schools. Teacher 08's students (new experimental course) not included.

[b] Grade 12 Regents Physics students of 17 teachers in 17 central New York State high schools.

TABLE 39

INTER-CORRELATION COEFFICIENTS OF ADJUSTED PHYSICS REGENTS SCORES AND OF ADJUSTED CPT AND TOUS PRE-POST GAINS[a]

			TOUS			
	RPE	CPT	I	II	III	Tot.
Physics Regents-June 1965 (RPE)	1 (216)	126 (186)	-184 (175)	056 (175)	-120 (175)	-022 (175)
Coop. Physics Test-raw scores (CPT)		1 (216)	-156 (190)	042 (190)	-180 (190)	050 (190)
TOUS-I-Science as an enterprise			1 (202)	364 (202)	184 (202)	669 (202)
TOUS-II-The scientist				1 (202)	232 (202)	686 (202)
TOUS-III-Methods and aims of science					1 (202)	669 (202)
TOUS-Total-Nature of science						1 (202)

[a] Pearson product-moment correlations. The upper number in non-diagonal entries is the correlation coefficient times 1,000; the lower number in all entries is the number of observations.

In terms of the Regents Physics Examination the results (Table 38) show the typical teacher to be effective for 75 per cent of course objectives.

The comparison of Cooperative Physics Test and Test On Understanding Science gains is striking. The teacher of the New York State Regents Physics Course is effective in producing a substantial student gain in the understanding of, and ability to use facts and principles of physics commonly associated with high school physics courses. He is not effective in producing a substantial student gain in an understanding of the nature of science in terms of science as an enterprise, the scientist as an individual, and the methods and aims of science.

CHAPTER EIGHT

TEACHER INFLUENCE AND EFFECTIVENESS

Introduction

Three major aspects of teacher effectiveness were treated in this study. Taken in terms of student achievement, the measures were class mean adjusted scores on Regents Physics Examination, June 1965 and class mean adjusted pre-post gain scores on Cooperative Physics Test and Test On Understanding Science-Total. The associated objectives of Regents Physics teaching are given in Chapter Seven. Score adjustments were a consequence of statistical control of student initial ability and previous achievement.

Tests for relations between influence and effectiveness, with the latter as dependent variable, were carried out. In each case the design called for a comparison of the effectiveness scores of teachers who were in the upper third on the influence variable to those who were in the lower third. Decisions were based on the Mann-Whitney U statistic.

The purpose of this chapter is to present the results of these analyses which represent the fulfillment of Objectives 2 and 3.

Variation of Effectiveness Scores

In view of the effectively random sampling necessary for Objective 1 of the study, there was no assurance beforehand that some of the teachers would be much more effective than others according to any one of the three effectiveness measures. The results of an analysis of the effectiveness data supported the original plan to compare the five "highest" teachers to the five "lowest" where "high" and "low" refer to the independent influence variables. The nature of the Mann-Whitney U test (to be discussed) and the variation of the effectiveness scores (dependent variable) were considered in this decision. The point is, there was ample variation in the various influence measures but substantially less variation in the effectiveness scores. The original plan, nevertheless, remained as the best method. The comparisons of effectiveness scores, although carried out after the U tests were performed, are better given as a preface to the U test results.

In comparing effectiveness scores, student's t test for comparing means was used where:

$$t = \frac{\overline{X} - \overline{Y}}{\left[\left(\frac{(N_X - 1)S_X^2 + (N_Y - 1)S_Y^2}{N_X + N_Y - 2}\right)\left(\frac{N_X + N_Y}{N_X N_Y}\right)\right]^{\frac{1}{2}}}$$

and $\overline{X}$ is the greater mean while S^2 is the unbiased variance estimate. Tables 40, 41, and 42 give the results of selected comparisons. All tests are one-tailed. In these tables "highest" and "lowest" refer to rankings on the particular effectiveness measure involved. Table 43 lists teachers' ranks on each of the effectiveness measures.

The purpose was to see if these class means were significantly different. For example, in Table 40 it is seen that fifth highest Regents Physics Examination adjusted class mean is not significantly greater, according to the t test, from the fifth lowest at the .10 level of confidence. Neither is the fourth higest significantly greater than the fourth lowest. However, the third highest is significantly greater than the third lowest at the .01 level of confidence. RPE class-wise standard deviations were such that it was clear that as one progressed to the highest and to the lowest, the differences of means were more extreme.

Effectiveness scores based on Cooperative Physics Test are somewhat less different than RPE scores.

Test On Understanding Science means are compared in Table 42. It is seen that whereas TOUS-I is suitable for analysis, TOUS-II and III are less useful. Note that for TOUS-III, even though the second highest and second lowest are significantly different at the .10 level, the highest is not significantly greater than the lowest. However, the variation in TOUS-Total is more "favorable" than any of the effectiveness criteria. The mean gains in TOUS-II and III were essentially nil in all the classes; not so with TOUS-I and TOUS-Total. These results should be kept in mind in interpreting the U test results to follow. Also, as a

TABLE 40

STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED SCORES ON PHYSICS REGENTS (JUNE 1965)

Means Compared	df	t	Significance
Fifth highest and fifth lowest	32	0.99	NS[a]
Fourth highest and fourth lowest	32	1.14	NS
Third highest and third lowest	36	2.69	Significant at .01 level

[a] Not significant at .10 level. All tests are one-tailed.

TABLE 41

STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON CPT

Means Compared	df	t	Significance
Fifth highest and fifth lowest	29	1.11	NS[a]
Fourth highest and fourth lowest	22	1.04	NS
Third highest and third lowest	17	1.40	Significant at .10 level
Second highest and second lowest	24	1.91	Significant at .05 level
Highest and lowest	23	3.50	Significant at .005 level

[a] Not significant at .10 level. All tests are one-tailed.

TABLE 42

STUDENT'S t TEST COMPARISONS OF SELECTED CLASS MEAN ADJUSTED PRE-POST GAIN SCORES ON TOUS-I, II, III AND TOTAL

Means Compared	df	t	Significance
TOUS-I			
Third highest and third lowest	15	1.21	NS[a]
Second highest and second lowest	24	2.96	Significant at .005 level
Highest and lowest	14	2.50	Significant at .025 level
TOUS-II			
Third highest and third lowest	23	1.40	Significant at .10 level
Second highest and second lowest	17	1.78	Significant at .05 level
Highest and lowest	16	1.83	Significant at .05 level
TOUS-III			
Fifth highest and fifth lowest	30	1.43	Significant at .10 level
Second highest and second lowest	22	1.48	Significant at .10 level
Highest and lowest	24	1.19	NS
TOUS-Total			
Fifth highest and second lowest	24	2.53	Significant at .01 level
Fifth highest and fifth lowest	29	2.11	Significant at .025 level

[a] Not significant at .10 level. All tests are one-tailed.

TABLE 43

RANKS ON CLASS MEAN ADJUSTED SCORES OF PHYSICS REGENTS AND OF CPT, TOUS-I AND TOUS-TOTAL PRE-POST GAINS[a]

Teacher	Physics Regents	CPT	TOUS-I	TOUS-Total
01	3.0	7.5	5.0	4.0
02	8.0	10.0	9.0	11.5
03	12.0	9.0	12.5	13.0
04	9.0	2.0	11.0	9.5
05	10.0	7.5	10.0	9.5
06	11.0	14.0	12.5	7.0
09	13.0	5.0	3.0	6.0
10	5.0	1.0	4.0	2.5
11	4.0	12.0	7.0	2.5
12	14.0	4.0	6.0	8.0
13	1.5	13.0	1.0	1.0
14	1.5	11.0	8.0	11.5
15	7.0	3.0	2.0	5.0
16	6.0	6.0	14.0	14.0

[a] Scores not available on all criteria for teachers 07, 08 and 17.

consequence of this, TOUS-II and III were deleted from the influence versus effectiveness comparisons.

The fact that TOUS-III measures student understandings concerning the methods and aims of science makes the result particularly startling. The implication that Regents Physics teachers are ineffective in producing a gain in such understandings is pronounced.

The Mann-Whitney U Tests

For a comparison of teachers who were "high" and "low" on any of the influence variables, the following conditions were considered:

1. The selection of a teacher for inclusion in the "high" group did not bias another's chances for inclusion. One teacher's score had no effect on what another's might be. The dependent effectiveness scores were not involved in the formation of "high" and "low" influence groups. By definition, the five highest ranking teachers were considered as "high" on the independent criterion and the five lowest ranking teachers were considered as "low." In this way the condition of independence of observations was realized.
2. There is no assurance that teachers' effectiveness scores are normally distributed in the population.

3. The variance of effectiveness scores of "high influence" teachers is not equal to the variance of effectiveness scores of "low influence" teachers.
4. There is no assurance that effectiveness scores are at least on an interval scale. They are surely at least ordinally scaled.

The purpose was to compare the effectiveness scores of two independent groups of teachers to see if the two groups have been drawn from the same population. The conditions of independence, normality, equality of variance, and interval scaling are necessary for use of the parametric Student's t test. When N is relatively large the assumption that the conditions are met is more sound than when N is small. Hence, it was reasonable to use Student's t in the foregoing section. However, with N equal to 5, the assumptions for parametric testing are placed in serious question. The assumptions associated with conditions 2, 3, and 4 above, were avoided by using the Mann-Whitney U test (non-parametric).

The U test provides an "exact statistic. U is given by:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

where R_1 is the sum of the ranks assigned to the group whose sample size is n_1. For all U tests in this study:

$$n_1 = n_2 - 5 .$$

Hence:

$$U = 40 - R_1 .$$

Thus, for each test, the effectiveness scores of teachers in the "high influence" group formed a set with the effectiveness scores of teachers "low" on the independent criterion. These ten scores were ranked with "1" assigned to the lowest effectiveness score and "10" assigned to the highest score. In testing the hypothesis that the effectiveness scores of "high influence" teachers are stochastically greater than those of "low influence" teachers, R_1 is the sum of the ranks of the effectiveness scores of "high influence" teachers.

The reader unfamiliar with U should bear two things in mind: (1) ranking is opposite to what is ordinarily done, i.e., "1" is assigned to the lowest score; and (2) the hypothesis of no difference between groups (H_0) is rejected with a low value of U.

One can test the opposite hypothesis, i.e., that effectiveness scores of "low" teachers are stochastically greater than those of "high" teachers by using:

$$U' = 40 - R_2 \qquad (n_1 = n_2 = 5) .$$

Here, R_2 is the sum of the ranks of effectiveness scores of "low influence" teachers. (See Siegel, 1956.) Finally, it is noted that:

$$U + U' = n_1 n_2 = 25 .$$

Directional hypothesis were of interest, hence all tests were one-tailed. An occasional tie in ranks was handled by assigning the mid-rank value to both tying elements. For example, two scores tied at rank 3 would both be assigned the value 3.5 and no score would be assigned a 3 nor a 4.

For all U tests reported the form of the hypothesis was:

H_0 : There is no difference between the effectiveness of teachers "high" on the independent criterion and the effectiveness of teachers "low" on the independent criterion.

The corresponding alternative hypothesis was:

H_1 : The effectiveness of "high" teachers is stochastically greater than that of teachers "low" on the independent criterion.

The sampling distribution under H_0 is known. Alpha was set at .05, hence it was decided to reject the null hypothesis if U were equal to or less than four. It is noted that under H_0, and with n_1 and n_2 equal to 5, the probability of U equals 5 is .075 and the probability of U equals 4 is .048.

Indirectness of Teacher Influence and Effectiveness

Mann-Whitney U tests were used to compare the effectiveness scores of "indirect" teachers with those of "direct" teachers. For these tests the "high influence" physics teachers were defined as the five teachers having the five highest scores on Total I/(I+D). Of the teachers studied, these teachers were most indirect in their influence in the classroom as determined by the interaction matrices. In an analogous fashion the five lowest on Total I/(I+D) were considered most direct.

Tables 44, 45, 46, and 47 give the results for each of the effectiveness criteria and respectively, for Total I/(I+D) based on Total Teaching, and Lecture, Laboratory, and Recitation Activities. Note that TOUS-II and III are omitted. It was decided to look at effectiveness in producing desirable student attitudes. Class mean scores on Student Questionnaire (SQ) provided attitude measures. This inclusion was purely a matter of additional interest aside from the major objectives of the study.

Demonstration Activity was omitted from the analysis because there were too few teachers with enough Demonstration time to facilitate sensible comparisons.

The tables give all pertinent information needed. Note that the statements of the particular H_0 and H_1 are given. Also, values of U' are provided as well as U values. In addition, ranges on the independent criteria are provided.

The results show that for each of the effectiveness criteria, there is no difference in effectiveness between physics teachers who are "more direct" and those who are "more indirect" on an overall basis (Table 44). However, teachers who exhibit more direct influence in the Lecture Activity (Table 45) are more effective in terms of Regents Physics Examination scores. They are also more effective

TABLE 44

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN TOTAL TEACHING[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	18	7	NS[b]
Class mean adjusted pre-post gain score on CPT	13	12	NS
Class mean adjusted pre-post gain score on TOUS-I	12	13	NS
Class mean adjusted pre-post gain score on TOUS-Total	14	11	NS
Class mean score on Student Questionnaire (attitude)	7	18	NS

[a] Ranges of Tot. I/(I+D) Percentage for Total Teaching: upper five teachers (indirect), 43.5--32.4; lower five teachers (direct), 24.7--10.6. U corresponds to the hypothesis H_1: The effectiveness of teachers who are indirect in total teaching is stochastically greater than that of direct teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

TABLE 45

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN LECTURE ACTIVITY[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	22	3	U' significant at .05 level
Class mean adjusted pre-post gain score on CPT	11	14	NS[b]
Class mean adjusted pre-post gain score on TOUS-I	21	4	U' significant at .05 level
Class mean adjusted pre-post gain score on TOUS-Total	23	2	U' significant at .05 level
Class mean score on Student Questionnaire (attitude)	11	14	NS

[a] Ranges of Tot. I/(I+D) Percentages for Lecture Activity: upper five teachers (indirect), 49.8--29.7; lower five teachers (direct), 19.4--3.0. U corresponds to the hypothesis H_1: The effectiveness of teachers who are indirect in lecture activity is stochastically greater than that of direct teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

TABLE 46

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN LABORATORY ACTIVITY[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	10	15	NS[b]
Class mean adjusted pre-post gain score on CPT	16	9	NS
Class mean adjusted pre-post gain score on TOUS-I	6	19	NS
Class mean adjusted pre-post gain score on TOUS-Total	8	17	NS
Class mean score on Student Questionnaire (attitude)	12	13	NS

[a] Ranges of Tot. I/(I+D) Percentage for Laboratory Activity: upper five teachers (indirect), 37.8--32.7; lower five teachers (direct), 24.2--18.6. U corresponds to the hypothesis H_1: The effectiveness of teachers who are indirect in laboratory activity is stochastically greater than that of direct teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

TABLE 47

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: INDIRECT INFLUENCE VERSUS DIRECT INFLUENCE IN RECITATION ACTIVITY[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	17	8	NS[b]
Class mean adjusted pre-post gain score on CPT	13	12	NS
Class mean adjusted pre-post gain score on TOUS-I	6	19	NS
Class mean adjusted pre-post gain score on TOUS-Total	8	17	NS
Class mean score on Student Questionnaire (attitude)	11	14	NS

[a] Ranges of Tot. I/(I+D) Percentages for Recitation Activity: upper five teachers (indirect), 53.8--38.0; lower five teachers (direct), 36.0--17.4. U corresponds to the hypothesis H_1: The effectiveness of teachers who are indirect in recitation activity is stochastically greater than that of direct teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

in producing understandings about the nature of science (TOUS).

Directness (or indirectness) in the laboratory (Table 46) is not a significant factor in teacher effectiveness according to the results of this study. It should be pointed out, however, that the interaction matrices from which the Total I/(I+D) scores were derived for this activity were based on only one period of laboratory observation in most cases. Hence the Total I-D percentage is substantially less reliable in this case. As a consequence, there is proportionately less conviction in the conclusion.

Finally, the test results concerning the Recitation Activity (Table 47) led to no significant results. That is, teachers who were "indirect" in this activity were no more effective than teachers who were "direct."

It is noteworthy that the results show a trend to favor indirect influence in Laboratory and Recitation as concerns effectiveness measured by Test On Understanding Science. The reported U value of 8 has a probability under the null hypothesis of 0.21.

Flexibility of Teacher Influence and Effectiveness

Three approaches were taken to produce measures that were considered likely associated with flexibility of teacher influence. The purpose of the following Mann-Whitney U tests was to investigate these flexibility measures as factors of teacher effectiveness. In each case the effectiveness scores of the five "most flexible" physics teachers were compared to those designated as the five "least flexible."

Flexibility score FSa was defined as the range of four I-D percentages, each of which was determined from the teacher's grand interaction matrix. The I-D percentages involved were Total I/(I+D), Revised I/(I+D), Rows 8 and 9 I/(I+D), and Extended I/(I+D). The greater the range of these I-D percentages, the more flexible is the teacher from the standpoint of interaction analysis.

Table 48 gives the results. Students of teachers who were more flexible according to FSa demonstrated more favorable attitudes as indicated by Student Questionnaire. All other FSa tests showed no difference in effectiveness.

Flexibility score FSb was defined as the standard deviation of Total I/(I+D) scores where the latter are based on grand matrices of each of the four Major Activities. In many cases teachers did not have four such matrices available. Matrices based on less than 600 tallies were not used as the corresponding Total I/(I+D) scores were felt to be insufficiently reliable for less than thirty minutes of observation time in an activity. The corresponding U test results are organized in Table 49. As far as this measure is concerned, the data tend to favor less flexibility of influence. This is opposed to the direction expected as a consequence of the research of Flanders at the junior high

TABLE 48

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSa[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	13	12	NS[b]
Class mean adjusted pre-post gain score on CPT	16	9	NS
Class mean adjusted pre-post gain score on TOUS-I	13	12	NS
Class mean adjusted pre-post gain score on TOUS-Total	7	18	NS
Class mean score on Student Questionnaire (attitude)	1	24	U significant at .01 level

[a] Flexibility index, FSa, is the range of four I-D percentages determined by a teacher's grand matrix. Ranges of FSa: upper five teachers ("more flexible"), 57.7--49.4; lower five teachers ("less flexible"), 35.2--21.2. U corresponds to the hypothesis H_1: The effectiveness of teachers who are "more flexible" is stochastically greater than that of "less flexible" teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

TABLE 49

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSb[a]

Criterion	U	U'	Significance
Class mean adjusted score on _Physics_ _Regents_ (June 1965)	9	16	NS[b]
Class mean adjusted pre-post gain score on _CPT_	25	0	U' significant at .005 level
Class mean adjusted pre-post gain score on _TOUS-I_	16	9	NS
Class mean adjusted pre-post gain score on _TOUS-Total_	16	9	NS
Class mean score on _Student_ _Questionnaire_ (attitude)	17	8	NS

---

[a] Flexibility index, FSb, is the standard deviation of Total I/(I+D) scores with the latter derived from a teacher's grand matrix on each of four Major Activities. Ranges of FSb: upper five teachers ("more flexible"), 18.0--11.8; lower five teachers ("less flexible"), 4.2--1.4. U corresponds to the hypothesis H_1: The effectiveness of teachers who are "more flexible" is stochastically greater than that of "less flexible" teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

school level. However, the only significant result is in the case of effectiveness as measured by Cooperative Physics Test. Here, U' equals zero, implying that physics teachers who were "less flexible" according to FSb were significantly more effective than teachers who were "more flexible." The result is significant at the .005 level.

A third measure, FSc, was devised and explored as a flexibility measure. FSc is defined as the number of standard deviations of chi-square where chi-square is derived by means of a Darwin Chi-Square test of homogeneity of a set of grand matrices. Each set was composed of a teacher's grand matrices from each of the four Major Activities. All matrices, regardless of the number of tallies, were used in this calculation (Wightman program on the 1604 computer). No significant results were found (Table 50).

On the whole, all efforts to isolate a flexibility factor related to effectiveness in Regents Physics teaching were unsuccessful.

Selected Interaction Matrix Measures and Effectiveness

A broad analysis was carried out to search for other aspects of the teacher's grand interaction matrix that might be related to one or all of the effectiveness measures. A number of I-D percentages and other cell and cell group percentages were identified as measures worthy of investigation. Non-interaction characteristics of the teachers and programs

TABLE 50

MANN-WHITNEY U TEST COMPARISONS OF REGENTS PHYSICS TEACHER EFFECTIVENESS SCORES: MORE FLEXIBLE VERSUS LESS FLEXIBLE AS INDICATED BY FSc[a]

Criterion	U	U'	Significance
Class mean adjusted score on Physics Regents (June 1965)	7	18	NS[b]
Class mean adjusted pre-post gain score on CPT	14	11	NS
Class mean adjusted pre-post gain score on TOUS-I	13	12	NS
Class mean adjusted pre-post gain score on TOUS-Total	17	8	NS
Class mean score on Student Questionnaire (attitude)	17	8	NS

[a] Flexibility index, FSc, is the standard deviations of chi square, the latter derived through a Darwin Chi Square test of homogeneity of a set of grand matrices. The set is composed of a teacher's grand matrix from each of the Major Activities, hence a maximum of four per set. Ranges of FSc: upper five teachers ("more flexible"), 16.4--11.2; lower five teachers ("less flexible"), 7.2--0.0. U corresponds to the hypothesis H_1: The effectiveness of teachers who are "more flexible" is stochastically greater than that of "less flexible" teachers. U' is the opposite hypothesis.

[b] Neither U nor U' are significant at the .05 level.

were examined as well. A 31 X 31 inter-correlation matrix of Kendall tau coefficients of correlation was produced using the RANKO Program (Searle, 1964c) on the 1604 computer.

The definitions of variables correlated are listed on the third page of Table 51. The table gives rows 1..7 across the inter-correlation matrix (i.e., columns 1..31).

Kendall's tau was used as the data met the condition that at least ordinal data was guaranteed on all variables. In addition, partial correlations were possible. It is noted that the Pearson r requires interval scaling and that Spearman rho will not give partial coefficients (see Siegel, 1956).

Various partial correlations and measures of concordance were produced in addition to the 31 X 31 inter-correlation matrix. However, only the correlations are reported as the latter two efforts led to no useful results.

Significantly correlating variables of Table 51 are of interest in themselves. (Perhaps non-significant correlations are equally interesting.) Many of the correlations "confirm" the results of U tests given earlier. Others of interest are somewhat peripheral to the major objectives of the study. For example, the teacher's preparation in physics was not significantly correlated with any of the effectiveness measures as seen by examining Column 29, Rows 1..7. The teacher's preparation in science was correlated .37 with student attitude. This correlation is the only one in Column 30 absolutely great enough to be significantly different from zero at the .05 level.

TABLE 51

KENDALL'S TAU CORRELATION COEFFICIENTS OF TEACHER EFFECTIVENESS SCORES WITH MEASURES FROM FLANDERS MATRICES AND OTHER TEACHER AND SCHOOL FACTORS[a]

Variable	1	2	3	4	5	6	7	8
1	1	-.20	.23	-.03	.24	.24	-.47[b]	-.15
2		1	.11	-.02	-.14	-.01	.28	.03
3			1	.39[b]	.55[c]	.60[c]	-.13	-.11
4				1	.47[b]	.55[c]	.14	-.04
5					1	.81[c]	-.05	-.27
6						1	.05	-.09
7							1	.19

Variable	9	10	11	12	13	14	15	16
1	-.29	-.04	.07	.11	-.28	-.20	-.16	.04
2	.23	-.21	-.41[b]	-.06	-.02	-.14	-.10	-.12
3	-.28	.00	-.09	-.15	.08	.11	.04	-.09
4	-.16	.13	-.18	-.20	.12	.13	.06	-.09
5	-.47[b]	.38[b]	-.11	-.27	.06	.16	.14	-.16
6	-.34[b]	.25	-.22	-.22	.06	.20	.18	-.09
7	.21	.21	-.12	-.19	.29	.43[b]	.38[b]	-.01

(continued)

TABLE 51 (continued)

Variable	17	18	19	20	21	22	23	24
1	-.13	-.01	-.08	.15	.63[c]	-.23	.00	.15
2	-.01	-.04	-.16	-.12	-.02	-.36[b]	.08	.36[b]
3	.09	.10	.06	.20	.37[b]	-.17	.11	-.06
4	.18	.11	.08	.11	-.15	-.03	-.13	.08
5	.31	.24	.04	.36[b]	.24	-.12	-.13	-.06
6	.22	.16	.12	.20	.26	-.24	.04	-.01
7	.08	-.02	.00	-.10	-.18	-.04	-.08	.23

Variable	25	26	27	28	29	30	31
1	-.13	.16	-.14	.14	.09	.09	-.21
2	.01	-.06	-.13	.04	.23	-.12	-.18
3	-.04	.07	-.30	.00	.13	-.04	-.32
4	.22	.07	-.26	-.22	.02	.09	-.24
5	-.02	.02	-.21	.02	.11	.13	-.30
6	.00	-.05	-.17	-.05	.02	.20	-.37[b]
7	.01	-.33	.07	.01	-.26	.37[b]	.02

[a] Number of observations is 14 for all correlations. Standard error for tau, based on the null case (that true tau is zero) is 0.20 for all correlations except the following, in which it is 0.21: 4-5, 4-6, 5-6, 4-24, 5-24, 6-24, and all correlations with variable 28. The following key identifies the variables:

(continued)

TABLE 51 (continued)

Variable Number	Identification
1.	Class mean adjusted score on Physics Regents, (June 1965).
2.	Class mean adjusted pre-post gain score on Cooperative Physics Test.
3.	Class mean adjusted pre-post gain score on TOUS-I-Science as an enterprise.
4.	Class mean adjusted pre-post gain score on TOUS-II-The scientist.
5.	Class mean adjusted pre-post gain score on TOUS-III-Methods and aims of science.
6.	Class mean adjusted pre-post gain score on TOUS-Total-Nature of Science.
7.	Class mean score on Student Questionnaire (attitude).
8.	Total I/(I+D)--from teacher's grand matrix of "total teaching" (I=cols.1..4; D=cols.5..7).
9.	Total I/(I+D) for Lecture Activity.
10.	FSa--flexibility index determined by the range of four I-D percentages from grand matrix of "total teaching."
11.	FSb--flexibility index determined by standard deviation of Total I/(I+D) scores from matrices of Major Activities.
12.	FSc--flexibility index determined by standard deviations of chi square from Darwin homogeniety test on Major Activities matrices.
13.	Rows 8 & 9 I/(I+D)--based on "total teaching" (I and D in rows 8 and 9, I=cols.1..4; D=cols.5..7).
14.	Revised I/(I+D)--based on "total teaching" (I=cols. 1..3; D=cols.6..7).
15.	Steady-State I/(I+D)--based on "total teaching" (diagonal cells--I=sum cells 1-1..3-3; D=sum cells 6-6..7-7).
16.	Per cent of student talk--based on "total teaching" (ratio cols. 8+9 to total tallies).
17.	Per cent content--based on "total teaching" (ratio cols. 4+5 to total tallies).
18.	Per cent steady lecture--based on "total teaching" (ratio 5-5 cell to total tallies).
19.	Per cent steady student talk--based on "total teaching" (ratio 8-8 cell + 9-9 cell to total tallies).
20.	Per cent lecture is of teacher talk--based on "total teaching" (ratio col. 5 to sum cols. 1..7).

(continued)

TABLE 51 (continued)

Variable Number	Identification
21.	Ratio student talk--initiation, to student talk--response (col. 9/col. 8).
22.	Row 9 I/(I+D) minus Row 8 I/(I+D)--based on "total teaching" (in row 9 or 8; I=cols.1..3; D=cols.6..7).
23	Per cent time in recitation activity (ratio of recitation time to total time less lab time).
24.	Per cent time in demonstration activity (ratio of demonstration time to total time less lab time).
25.	Enrollment, grades 10-12.
26.	Weekly student time (minutes of scheduled instruction per week per student).
27.	Per cent total physics enrollment in school is of grades 10-12 enrollment.
28.	Teacher's experience in secondary science (years).
29.	Teacher's preparation in physics (credit hours of college-level physics).
30.	Teacher's preparation in science (credit hours of college-level science; includes biology, chemistry, physics, etc.).
31.	Teacher load--assigned (hours per week of assigned duties; includes study-hall, homeroom, etc.).

Data for teachers 07, 08 and 17 were eliminated in the determinations of correlations. Teacher 08's course was based on an experimental Regents physics syllabus; there were no measures on variables one through six for teachers 07 and 17 in view of a majority enrollment of 11th grade students in physics.

[b] Significantly greater or less (depending on whether tau is + or -) than zero at the .05 level.

[c] Significantly greater or less (depending on whether tau is + or -) than zero at the .01 level.

It is especially noteworthy that the ratio of the amount of student initiated talk to student response, i.e., Column 9 divided by Column 8 in the interaction matrix, has a .63 correlation with effectiveness as measured by Regents Physics Examination. This is significantly greater than zero at the .01 level. The same ratio is significantly related to effectiveness as measured by TOUS-I.

One can glean a lot of general information by a careful examination of the table. The first seven rows of the matrix were presented in lieu of a detailed discussion of the correlations.

The correlations served as a guide for additional U test comparisons. U test results that were significant are presented in Table 52. Ranges of the variables involved are given to support the information.

It is noted that teachers who were more direct according to the Rows 8 and 9 I-D percentage were more effective in terms of RPE. This result, opposite to the direction expected, may be a consequence of the subject matter orientation of RPE. That is, the physics teachers and the examination are especially concerned with student understanding of the basic facts and principles outlined in the minimum required syllabus of the Regents Physics Course. Perhaps teachers who were more prone to criticize and reject student statements placed more emphasis on the Regents topics as a consequence. This can only be a speculation.

TABLE 52

MANN-WHITNEY U TEST COMPARISONS OF TEACHER EFFECTIVENESS SCORES USING HIGH VERSUS LOW GROUPS ACCORDING TO SELECTED CORRELATED VARIABLES[a]

Effectiveness Score	High-Low Criterion[b]	U	U'	Significance
Class mean adjusted score on *Physics Regents* (June 1965)	Rows 8 & 9 I/(I+D)--on total teaching	23	2	U' significant at .05 level
Class mean adjusted score on *Physics Regents* (June 1965)	Col. 9/Col. 8 (S initiation-response ratio	2	23	U significant at .05 level
Class mean adjusted pre-post gain score on *TOUS-I*	% total physics enrollment is of grades 10-12	21	4	U' significant at .05 level
Class mean adjusted pre-post gain score on *TOUS-III*	Total I/(I+D) for lecture activity	23	2	U' significant at .05 level
Class mean adjusted pre-post gain score on *TOUS-III*	Col. 9/Col. 8 (S initiation-response ratio)	4	21	U significant at .05 level
Class mean adjusted pre-post gain score on *TOUS-III*	Teacher load-assigned	23	2	U' significant at .05 level

[a] U corresponds to the hypothesis H_1: The effectiveness of teachers who are "high" on the independent variable is stochastically greater than that of teachers who are "low" on the independent variable. U' is the opposite hypothesis.

[b] "High" ranges and "low" ranges are as follows: Rows 8 & 9 I/(I+D), 70.4%--59.8% and 53.7%--33.8%; Col. 9/Col. 8, 6.5--3.6 and 2.4--1.1; % total physics enrollment is of grades 10-12, 11.6%--8.4% and 7.0%--5.7%; total I/(I+D) for lecture activity, 49.8%--29.7% and 19.4%--3.0%; and teacher load-assigned, 28.9--23.5 hours per week and 21.8--18.4 hours per week. In all cases, for both the "high" group and the "low" group, N = 5.

The student initiation-response ratio shows up equally well as a factor of effectiveness as measured by Regents Physics Examination. The result shows that in classes where there were relatively more 9's than 8's (Flanders' categories), the students attained higher achievement on the Regents Physics Examination. This implies that teachers who ask questions that are more thought provoking are more effective on RPE.

The two results seem to oppose each other at first appearance. One possible resolution is the speculation that physics teachers who more often solicit student ideas through thought provoking questions, but then criticize these ideas and responses, i.e., respond directly to them, are more effective in terms of RPE.

The results concerning TOUS-III in Table 52 are not to be taken very seriously in light of the fact that gains on TOUS-III were not very different from class to class. The results were presented as a matter of completeness.

Concluding Remarks

The broad analyses of the search for factors of teacher-student verbal interaction related to teacher effectiveness led to results that might appear disappointing. Although some factors were isolated in connection with some of the effectiveness measures, no one factor appeared as a variable highly related to teacher effectiveness in general. However, the findings do provide evidence that will be useful in guiding a future continuation of the search for more information about desirable behavioral aspects of the teaching process in high school physics.

CHAPTER NINE

IMPLICATIONS OF MAJOR FINDINGS

The purpose of this chapter is to present seemingly pertinent implications derived from the major findings of the study. The data and conclusions have been set forth in earlier chapters. Although the conclusions were reached through essentially objective means, some of the corresponding interpretations and implications are necessarily the consequence of relatively subjective considerations.

Implications Concerning the Usefulness of Interaction Analysis

The Flanders method of interaction analysis has been shown, in this study, to be a feasible method of acquiring useful data concerning what takes place in the high school physics classroom. The interaction matrices provided extensive reliable and objective quantitative information on aspects of teacher-student verbal interaction along a directness dimension of teacher influence.

Techniques were developed to make it possible for a single observer to effectively capture the transpired discourse of physics classrooms. In this connection a set of special ground rules was designed to clarify Flanders' category definitions as they apply to the intricacies of physics teaching. By employing a running commentary to supplement tape recordings, it was shown that the discourse could be stored and reliably coded in the investigator's laboratory. It appears that an investigator can study the same teaching in different ways if the methods of this study are used.

Numerous matrices were compared in terms of the Scott Pi reliability coefficient and Darwin Chi-Square. Pi is a function of the column (or row) totals; the Darwin statistic compares matrices in terms of tally distributions among cells. Pi is generally accepted as a good measure of observer reliability. In some cases Pi indicated two matrices were alike while chi-square indicated the two were different. Some explanation of this discrepancy was possible in terms of the proportion of cells having zero tallies. However, there is a need to learn more about how the two measures relate as the data raise questions about the validity of Pi as a reliability indicator.

It was found that a physics teacher's distribution of tallies in the matrix is practically constant providing all observations are taken in a particular type of activity (lecture, laboratory, recitation-discussion). This is not true, however, when pairs of composite matrices representing all activities are compared. Two important implications result:

1. A static tally distribution is associated with consistencies of teacher influence. Hence, the interaction matrix provides reliable quantitative data

on teacher style in physics classes. An ability to quantitatively measure teacher style facilitates possibilities of experimenting with styles and developing styles of prospective physics teachers.

2. Teachers' styles of teaching vary from one Major Activity to another. This suggests the need for consideration of the kind of activity involved when investigators concern themselves with teacher style.

Implications Concerning Directness of Teacher Influence

The Normative Data

Normative data on teacher behavior in high school physics classes was provided in this study through the production of a grand composite matrix. This interaction matrix contains over 86,000 tallies and is based on more than seventy-two hours of coded teacher-student interaction. In addition, grand composite matrices were produced for each of four Major Activities. Analyses of the composite matrices provided a reliable picture of what takes place in the physics classroom when physics teaching is viewed through the Flanders categories. Comparisons to normative data on junior high school teachers (Flanders' studies of seventh-grade mathematics and eighth-grade social studies teachers) were made. Results are detailed in Chapter Six. In what follows salient features are reviewed and associated speculations are attached.

1. Physics teachers provide 82 per cent of the verbal communication in the classroom. Of teacher talk 73 per cent is associated with direct influence. The corresponding junior high school figures are both about 67 per cent. Lecturing accounts for 92 per cent of teacher directness in physics classes. Physics teachers do relatively more lecturing but junior high teachers do more directing and criticizing. In physics classes 16 per cent of teacher talk involves questioning; half as frequently teacher talk is associated with acceptance and clarification of student ideas. Compared to junior high teachers, physics teachers use more praise and clarification and these are more likely to appear in a sustained sense. Hence, although little usage is made of acceptance and clarification, the results imply that physics teachers make better usage of these.

Overall, physics teachers exhibit about as much indirect influence as do junior high teachers, i.e., Total I/(I+D) is about the same (27%). However, when emphasis on content is removed from consideration (Revised I/(I+D), physics teachers are markedly more indirect (65% compared to 40%). Although more indirect, they are still not very indirect on the Revised I-D measure.

It is seen that although both physics teachers and junior high teachers are equally direct in their influence, they are direct and indirect in different ways. That is, there are marked differences in the way tallies are distributed among the direct categories and among the indirect categories. These differences point to the need of normative data for various subject areas if one attaches value to the kinds of information extractable from the Flanders matrix.

2. The finding that physics teachers do more talking, and in this sense more lecturing, than junior high teachers is connected with less student participation. On the other hand, more of the student talk in physics classes is student-initiated talk. Also, there is much less sustained talk of the limited response type (in the 8-8 cell, 0.4 per cent of the matrix compared to 2.9 per cent for eighth-grade mathematics). Thus, the kinds of student talk differ between the two levels. This supports the implication of Item 1 above.

The contention that student learning is enhanced through participation is generally accepted. The possibility of improving the amount and type of student participation in physics classrooms through interaction analysis is implied. Special categories to study types of teacher statements that solicit student response (e.g., teacher questions) and types of response produced could be developed in a way appropriate to the physics curriculum. Prospective teachers could practice getting and using student participation; interaction analysis would provide the feedback data. Such efforts might prove very fruitful in the practice teaching situation.

3. For physics teachers, 1.5 per cent of teacher statements are associated with praise and encouragement, and less than 0.3 per cent belongs to Category 1, Acceptance of Feeling. The employment of social skills connected with aspects of positive motivation is thus an uncommon phenomenon in the high school physics classroom. If a greater employment of such skills is desirable in work with groups of young people, then interaction analysis might well be employed in science teacher training programs in developing these skills in prospective and in-service teachers.

4. The Column 9 to Column 8 ratio is almost six times greater for physics teachers than Flanders' results showed for junior high school (2.8 and 0.5). This implies that physics teachers ask more thought provoking questions and get more student initiation. One assumption of a recent study in general science classrooms was that "the kinds of questions teachers ask are an indication of the quality of teaching that is going on and the levels of thinking that are being stimulated . . ." (Kleinman, 1965). A study focused on teacher questions in physics classes is recommended.

5. The results of this study facilitate an implication that physics teachers are not building on student ideas. Several findings support this contention: (a) The physics teacher is substantially more accepting toward student statements that he has solicited than he is of student initiated statements. This is considered striking in view of

a marked lack of social misbehavior on the part of students. (b) Only 8.5 per cent of teacher statements are associated with acceptance and clarification. (c) There is twice as much sustained student initiation (9-9 cell) as there is sustained acceptance and use of student ideas (3-3 cell, 2 per cent of total). Perhaps teachers could practice and improve their ability to build on student ideas as they develop concepts in the classroom.

6. The results of this study indicate that much of physics teaching is "telling." The normative data indicates relatively little learning through discovery. This is in sharp contrast to what educational psychologists and science educators would like to have. A point of Piaget, transmitted by Duckworth, is illustrative:

> You cannot further understanding in a child simply by talking to him. Good pedagogy must involve presenting the child with situations in which he himself experiments, in the broadest sense of that term--trying things out to see what happens, manipulating symbols, posing questions and seeking his own answers, reconciling what he finds one time with what he finds at another, comparing his findings with those of other children. (Piaget, according to Duckworth, 1964.)

Such views are amplified in the report of a recent conference on cognitive studies held at Cornell University (see Ripple and Rockcastle, editors, 1964).

A major attack on science curriculums from kindergarten through grade twelve is underway. Some progress has already been made. Johnson's view of "what is likely to become more noticeable in the years ahead" are quoted as follows:

> Change from one problem-solving method to many relatively unstructured methods of inquiry. . . .
>
> Change from an emphasis on accumulating knowledge to an emphasis on how to find out and create knowledge. . . .
>
> Change from facts and factual concepts as instructional goal to skills in inquiry as the teaching goals. . . .
>
> Change from teacher-selected concepts as instructional goals to concepts as they may arise in the process of confirming or rejecting hypotheses. . . .
>
> Change from science as something to be learned from books to science as something that grows out of a series of experiments. . . . (1963b)

The curriculum changes being effected reflect a concern for new and exciting objectives. It is clear that what takes place in the classroom will have to be examined in terms of these objectives. What are the prospects for new programs in physics classrooms if physics teaching is "telling"? Can you "tell" a student how to inquire?

The Laboratory Activity

It is a contention of the author that the laboratory activity may well be the phase of physics teaching most seriously in need of change. In Regents Physics laboratory programs students verify and confirm subject matter facts and principles already presented by the teacher in the classroom. It is the author's impression that these are things the teacher feels are important for the student to know. More often than not they are things that the teacher feels the student will be tested on when the final examination is given. A striking implication of the data was that there is little or no teaching to cultivate student inquiry in the high school physics laboratory. It appears that students have little or no opportunity to discover anything. The following findings are pertinent:

1. Nineteen per cent of teacher statements in the laboratory are directions and commands to which students are to comply. This is six times more directions than in any of the other three Major Activities.
2. There is less teacher praise and encouragement in the laboratory than in any other Major Activity.
3. There is less teacher acceptance and use of student ideas in the laboratory than in any other Major Activity.
4. There is more teacher criticism in the laboratory than in any other Major Activity.
5. Although the frequency of student initiated responses is greatest in the laboratory, there are twice as many teacher statements of his own ideas and opinions.
6. The teacher is much more direct on the Extended Influence variable in the laboratory than he is in any of the other Major Activities. Extended I/(I+D), where I is extended indirect influence and D is extended direct influence, is 9.7 per cent for Laboratory. The next lowest value is 63.3 per cent (Recitation Activity).

7. In the Laboratory Activity, solicited student statements, i.e., responses to "narrow" questions, are more readily accepted by teachers than are student initiated responses.

The problem of helping teachers to promote inquiry in their laboratory programs is one of prime importance that must be attended to. Interaction analysis and the data of this study can assist research directed to this problem.

Implications Concerning Teacher Effectiveness

No single measure of this study appeared as a factor of teacher effectiveness for all of the three aspects of effectiveness measured. The results of all tests of effectiveness are given in Chapter Eight. The failure to turn out a major factor is not surprising. It is common with the consensus of other research efforts on effectiveness.

It should be noted that extensive linear multiple regression analyses were carried out to equate students on various control factors of initial ability and previous achievement. Hence the failure to isolate a significant effectiveness factor points up the maze of factors that play a part in affecting student achievement.

Several correlates of particular teacher effectiveness measures were found, however. The ratio of student initiation to student response (Column 9 to Column 8) had a Kendall tau correlation of 0.63 with class mean adjusted Regents Physics Examination scores. This is significantly greater than zero at the 0.01 level of confidence. Also, this ratio was significantly correlated with class mean adjusted pre-post gain scores on Test On Understanding Science-I (.05 level). The corresponding Mann-Whitney U test showed that the five physics teachers who were highest on the ratio had significantly higher class mean adjusted RPE scores than the five lowest (U = 2, .05 level). In contrast to this result the U test showed that lower scores on Rows 8 and 9 I/(I+D) corresponded to significantly greater effectiveness based on RPE. A rationalization suggesting the more effective teacher to be more soliciting and yet more critical in his response to student remarks was presented. Such a view is, at best, "rough speculation." What is likely of value here is the support given to the need for concern about student participation in the classroom.

Three measures of flexibility of teacher influence were explored. None showed up as factors of teacher effectiveness according to the student achievement measures used. In one case, however, greater flexibility was significantly associated with higher scores on an attitude inventory. Thus, "more flexible" corresponded to "better attitudes." At best, these results do little more than indicate the need for further research. The failure to relate flexibility to student achievement in high school physics is in sharp contrast to the speculations of Flanders.

As it turned out the real finding of this phase of the study was the fact there were essentially no student gains over a five month period on Test On Understanding Science. This test, reliable and valid, measures understandings of the nature of science in terms of science as an enterprise, the scientist, and the methods and aims of science. Gain in understanding of the nature of science, in this sense, is viewed by many as an important objective of science teaching (c.f., Johnson, 1963b). The implication is that these views are not effectively shared by high school physics teachers. Physics teachers are primarily concerned for student performance on subject matter examinations. The kind of teaching that transpires in the laboratory is another consequence of this concern.

The recent comparison made by Trent (1965) of PSSC (Physical Science Study Committee) and traditional physics courses in terms of attainment of science understanding in California schools, as measured by Test On Understanding Science, showed no significant differences. Prior science understanding and mental ability were statistically controlled. Trent suggested the mood and style of the teacher as a factor possibly more important for this objective, than curriculum content. The results of this study show that, for Regents Physics classes, the "mood and style" is not a factor in TOUS gain. There is no TOUS gain in the average class.

If understanding of the nature of science is a desirable outcome of Regents Physics teaching, then the implication is for a need to include some specification of these understandings in the Regents Physics syllabus. Both the current Regents syllabus and the new syllabus (to be adopted throughout the State in 1967) are limited to a specification of subject matter topics, principles, and facts. That is, the teacher is guided in terms of what physics subject matter to cover (see Van Hooft, 1964). The new course is personally viewed as a better course in these terms. It is likely that better laboratory programs will also be a consequence. This may help, but is it enough? Listing subject matter topics to be covered is only part of the job of curriculum development. There is a need to specify sound objectives of physics teaching; sound objectives should include objectives concerning the nature of science.

Implications Concerning Further Research

Recommendations of further research, stemming from the major findings of this study, are embraced within the foregoing discussion of this chapter. The findings and rationale upon which the recommendations are based were given. The following list is a summary of the research implications. The order is not indicative of relative importance; some of the items are interrelated. There is need to:

1. Use the stored classroom discourse, acquired in this study, to analyze high school physics teaching in different ways.
2. Learn how the Scott Pi and Darwin Chi-Square statistics relate as indicators of observer reliability.
3. Experiment with various styles of physics teaching as measured by Flanders interaction matrices.
4. Develop and employ special categories for an analysis of the types of teacher statements that are more effective in fostering student initiated response and analyze teacher reaction to these kinds of student-teacher communication.
5. Explore ways in which prospective teachers can learn to employ social skills known to be valuable in work with groups.
6. Develop and use a set of categories focusing on the kinds of questions teachers ask in physics classrooms.
7. Study ways in which physics teachers can improve their ability to build on student ideas.
8. Study methods of teaching a teacher how to teach through inquiry. Of special importance is the need to help physics teacher promote inquiry in their laboratory programs.
9. Produce and analyze extensive normative interaction data on physics teaching in the Laboratory Activity.
10. Investigate further the implication that the more

effective physics teacher is both more soliciting of student participation and more critical in his responses to student remarks.

11. Continue the search for useful measures of flexibility of influence in physics classrooms.
12. Develop means of encouraging attainment of objectives concerning student understanding of the nature of science.

Two additional research needs, not as yet mentioned, appear as problems of primary importance: (1) The analysis of patterns of tally distributions within the Flanders matrix gives promise of a means of acquiring extended information concerning what actually takes place in high school physics classrooms. The notion of "check-pointing," discussed in Chapter Six, is an example. (2) Written transcripts can be produced from the tape recordings of this study. Teaching can be analyzed in terms of physics content, i.e., in terms of the physics that was communicated. Such a study could supplement the current study which employs type of teacher statement as the unit of analysis.

CHAPTER TEN

SUMMARY

Objectives

The expanding interest in research on teaching is motivated by the recent cultivation of various techniques for systematic observation of teacher and student behavior in the classroom. One of the most promising approaches is Flanders' method of interaction analysis. Based on the dimension directness of teacher influence and facilitating an objective analysis of teacher behavior, the Flanders technique was employed in an analytical study of high school physics teaching. The objectives were:

1. To describe physics teaching in terms of teacher-student verbal interaction using the Flanders categories.
2. To determine relationships between teacher effectiveness and directness of teacher influence.
3. To determine relationships between teacher effectiveness and flexibility of teacher influence.
4. To produce implications for (a) supporting or challenging the findings of others, (b) ways in which the education and supervision of science teachers can be improved, and (c) further research on teaching secondary science.

5. To provide a store of useful data on the discourse of high school physics teaching.

Procedures

A sample of seventeen high school physics teachers of the New York State Regents Physics Course and their students were studied. Seventeen central New York State high schools were involved. According to Kolmogorov-Smirnov tests of twelve pre-selected characteristics of teachers and schools, the sample adequately represented a defined population of 131 public high schools in fourteen counties surrounding Ithaca, New York. It was assumed that (1) the Regents Physics Course is typical of high school physics curricula, and (2) verbal behavior in the classroom is an adequate sample of total teacher behavior.

Classes were observed for four consecutive days in the fall and for three consecutive days in the spring (1964-1965). Additional observations were necessary to include laboratory periods; the first observation served acclimation purposes. Over 120 periods of teaching provided the data. In each case, Flanders' categories were applied to the classroom discourse taken from sound recordings supplemented by written running commentaries. The observations focused on the teacher and the teacher-student verbal interaction that transpired.

A grand composite Flanders matrix was produced and analyzed. This matrix was based on seventy-two hours of interaction and contained over 86,000 tallies. By examining distributions of tallies in the ten columns (one for each category) and in various combinations of the 100 matrix cells, quantitative data on physics teaching were obtained. These normative data facilitated a reliable description of what takes place in the physics classroom in terms of the directness of teacher influence dimension. In addition, the teaching was segmented into four Major Activities--Planned Demonstration, Lecture, Laboratory, and Recitation-Discussion. A composite matrix for each Major Activity was similarly produced and analyzed.

An assessment of student attitude using an adaptation of the Michigan Student Questionnaire enhanced the interaction analyses. Also, a Teacher Questionnaire was devised to estimate the effect of the observer in the classroom.

Fall and spring matrices were compared for each teacher using the Darwin Chi-Square Test to provide an indication of how well the tally distributions represented teacher consistencies of behavior. Such consistencies were considered as aspects of teacher style of teaching.

Three aspects of teacher effectiveness (dependent variables) were assessed in terms of corresponding class achievement measures: (1) class mean adjusted scores on Regents Physics Examination, June 1965; (2) class mean adjusted pre-post gain scores on Cooperative Physics Test;

and (3) class mean adjusted pre-post gain scores on Test On Understanding Science. Adjusted gain scores were given by the post-test minus the adjusted pre-test scores. Scores were adjusted by means of linear multiple regression analyses with aptitude and previous achievement measures as independent variables.

Indirectness and flexibility of teacher influence measures were taken from each teacher's composite matrix for Total Teaching and for each of the Major Activities. Teachers were grouped as "high" and "low" on these independent criteria. For each aspect of effectiveness, effectiveness scores of the upper third teachers ("high," $N = 5$) were compared to those of the lower third ("low," $N = 5$) using the Mann-Whitney U Test. The null hypothesis was "H_0 : There is no difference between the effectiveness of teachers 'high' on the independent criterion and the effectiveness of teachers 'low' on the independent criterion." The corresponding alternative hypothesis was "H_1 : The effectiveness of 'high' teachers is stochastically greater than that of teachers 'low' on the independent criterion." All tests were one-tailed; alpha equalled 0.05.

A 31 X 31 Kendall tau correlation matrix was produced to extend the search for correlating variables and correlates connecting effectiveness with aspects of the Flanders matrix and with various teacher and school characteristics.

It should be noted that all plots of Flanders matrices and calculations of interaction matrix measures, regressions, and correlations were carried out using the Control Data 1604 computer at Cornell University.

Major Findings

Techniques

The results showed that it was possible for a single observer to effectively capture and reliably analyze the transpired discourse of the high school physics classroom. It was found that a teacher's matrix is practically constant providing all observations are taken in a particular Major Activity. This implied that matrix measures were, in effect, measures of teacher style and that teacher style varied from one Major Activity to another.

Teacher Influence in the Classroom

Results of analyses of composite interaction matrices led to the following general conclusions:

1. Overall, physics teachers are direct in their influence; 73 per cent of all teacher statements were in direct influence categories. This agrees with Flanders' findings for junior high school teachers. When emphasis on content is removed from consideration, physics teachers are markedly more indirect. Compared to junior high teachers, physics teachers use more sustained praise and

clarification. They do more lecturing and less criticizing and directing. Marked differences between physics and junior high supported the original speculation of a need for normative data for different subject areas.

2. Physics teacher talk is distributed as follows: Acceptance of Feeling, 0.3 per cent; Praise or Encouragement, 1.5 per cent; Accepts or Uses Student Idea(s), 8.3 per cent; Asks Questions, 16.3 per cent; Lectures, 67.6 per cent; Gives Directions, 3.3 per cent; and Criticizes, 0.9 per cent.
3. Compared to junior high, there is less student participation in physics classes; however, physics students bring more of their own ideas into discussions (greater by a factor of six). An implication is that physics teachers ask more thought provoking questions.
4. The employment of social skills connected with aspects of positive motivation is relatively uncommon in the high school physics classroom (although more common there than in junior high).
5. Much of physics teaching is "telling." The matrix data indicates relatively little learning through discovery. Teachers rarely build on students ideas in developing concepts. The implication is that there is little student inquiry in the learning process.

Teacher Influence in the Laboratory Activity

Composite matrices of the Major Activites were compared. The results showed sharp discrepancies for Laboratory. The implication was that in the laboratory there is little or no teaching,to cultivate inquiry or to facilitate discovery. Teacher influence in Laboratory compares to that of the other Major Activites as follows: (1) six times more directions, (2) less praise and encouragement, (3) less acceptance and use of student ideas, (4) more criticism, and (5) much more direct influence associated with affective aspects of motivation. An implication of prime importance is that teachers do little promotion of inquiry in their laboratory programs.

Teacher Influence and Effectiveness

Directness of influence and three measures of flexibility were explored. No single measure of the study appeared as a factor of teacher effectiveness for all aspects of effectiveness considered. Non-null results were limited to the finding of several significant correlates of some effectiveness measures.

A major finding was that class mean adjusted gains on Test On Understanding Science were approximately zero. Hence, student gains were found to be limited to subject matter facts and concepts. Since understandings of the nature of science (process, methods, aims, etc.) are considered important, the lack of gain on this test raises serious questions about the curriculum and physics teaching.

Further Research

The study yielded various implications for further research. In particular the discourse acquired can be analyzed in various ways. An analysis of the physics content of teacher statements, using typewritten transcripts of the tape recordings might well supplement this study. Also, special categories focusing on such things as kinds of teacher questions and types of teacher statements that foster student initiated response could be profitably used. Attempts should be made to study ways in which prospective physics teachers could be taught to teach through inquiry, to build on student ideas, and to better employ social skills thought to be valuable in work with groups. Also, additional studies to facilitate improved physics laboratory programs are needed. Research on secondary physics teaching must be extended along many paths to foster the realization of a continually improved education in science for our young people.

LITERATURE CITED

1. ALLEN, DWIGHT W., and JIMMIE C. FORTUNE. "An Analysis of Micro-Teaching: A New Procedure in Teacher Education." Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1965.

2. AMIDON, EDMUND. "The Observational Technique of Interaction Analysis Applied to the Classroom: Procedures and Limitations." Paper presented at the 1963 convention of the American Educational Research Association, Chicago, 1963.

3. AMIDON, EDMUND J., and NED A. FLANDERS. "The Effects of Direct and Indirect Influence on Dependent-Prone Students Learning Geometry," _Studies of Classroom Interaction and Influence_. Research Memorandum BER-61-3. Minneapolis: University of Minnesota, College of Education, Bureau of Educational Research, May 1961.

4. ______, and NED A. FLANDERS. _The Role of the Teacher in the Classroom: A Manual for Understanding and Improving Teachers' Classroom Behavior_. Minneapolis: Paul S. Amidon and Associates, Inc., 1963.

5. ______, and ANITA SIMON. "Implications for Teacher Education of Interaction Analysis Research in Student Teaching." Paper presented at the thirty-eighth annual meeting of the American Educational Research Association, Chicago, February 1965.

6. ANDERSON, H. H., and HELEN M. BREWER. "Studies of Teachers' Classroom Personalities. II. Dominative and Socially Integrative Behavior of Kindergarten Teachers," _Psychological Monographs No. 6_, 1945.

7. ASCHNER, MARY J. McCUE. "The Analysis of Verbal Interaction in the Classroom," in _Theory and Research in Teaching_, Arno A. Bellack, editor. New York: Columbia University, Teachers College, Bureau of Publications, 1963.

8. ______, _et al_. _A System for Classifying Thought Processes in the Context of Classroom Verbal Interaction_. Urbana, Illinois: University of Illinois, Institute for Research on Exceptional Children, 1962.

9. BALES, R. F. *Interaction Process Analysis*. Cambridge, Massachusetts: Addison-Wesley, 1950.

10. BARR, A. S. *Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies*. Bloomington, Illinois: Public School Publishing Company, 1929.

11. ______. "The Measurement of Teaching Ability," *Journal of Educational Research*, 28:561-569, 1935.

12. ______, *et al*. *The Measurement and Prediction of Teacher Effectiveness*. Madison, Wisconsin: Dembar Educational Research Service, 1961.

13. BELLACK, ARNO A., and J. R. DAVITZ. *The Language of the Classroom*. New York: Columbia University, Teachers College, Institute of Psychological Research, 1963.

14. BLUMBERG, ARTHUR, and EDMUND AMIDON. "Teacher Perceptions of Supervisor-Teacher Interaction." Paper presented at the 1965 convention of the American Educational Research Association, Chicago, 1965.

15. BRAKKEN, EARL. "An Analysis of Some of the Intellectual Factors Operative in PSSC and Conventional Physics." Paper presented at the thirty-eighth annual meeting of the National Association for Research in Science Teaching, Chicago, 1965.

16. BURKE, PAUL J. *Cooperative Physics Test Form Z*. Princeton, New Jersey: Educational Testing Service, Cooperative Test Division, 1950.

17. COGAN, MORRIS L. "Theory and Design of a Study of Teacher-Pupil Interaction," *Harvard Educational Review*, 26:315-42, 1956.

18. ______. "The Behavior of Teachers and the Productive Behavior of Their Pupils: I. 'Perception Analysis,' and II. 'Trait Analysis,'" *Journal of Experimental Education*, 27:89-124, December 1958.

19. ______. "Research on the Behavior of Teachers: A New Phase," *Journal of Teacher Education*, 14:238-43, September 1963.

20. COOLEY, W. W., and L. E. KLOPFER. *TOUS, Test on Understanding Science, Form W*. Manual for Administering, Scoring, and Interpreting Scores. Princeton, New Jersey: Educational Testing Service, 1961.

21. CORNELL, F. G., C. M. LINDVALL, and J. L. SAUPE. *An Exploratory Measurement of Individualities of Schools and Classrooms*. Urbana, Illinois: University of Illinois, Bureau of Educational Research, 1952.

22. DARWIN, J. H. "Note on the Comparison of Several Realizations of a Markoff Chain," *Biometrika*, 46:412-19, 1959.

23. DAY, WILLIAM W. "Physics and Critical Thinking: A Comparison of PSSC and Traditional Physics." Paper presented at the 1965 convention of the National Association for Research on Science Teaching, Chicago, 1965.

24. DuBOIS, PHILIP H. *Multivariate Correlational Analysis*. New York: Harper, 1957.

25. DUCKWORTH, ELEANOR. "Piaget Rediscovered," *Piaget Rediscovered*, Richard E. Ripple and Verne N. Rockcastle, editors. Ithaca, New York: Cornell University, School of Education, 1964.

26. FLANDERS, NED A. "Personal-Social Anxiety as a Factor in Experimental Learning Situations," *Journal of Educational Research*, 45:100-10, 1951.

27. ______. *Teacher Influence, Pupil Attitudes and Achievement*. U. S. Office of Education Cooperative Research Project No. 397. Minneapolis: University of Minnesota, 1960.

28. ______. "Intent, Action and Feedback: A Preparation for Teaching," *Journal of Teacher Education*, 14:251-60, 1963a.

29. ______. "Helping Teachers Change Their Behavior." Projects Number 1721012 and 7-32-0560-171.0, U. S. Office of Education. Ann Arbor, Michigan: University of Michigan, School of Education, 1963b.

30. ______. *Interaction Analysis in the Classroom: A Manual for Observers*. Revised edition. Ann Arbor, Michigan: University of Michigan, School of Education, 1964a.

31. ______. "Some Relations Between Teacher Influence, Pupil Attitudes, and Achievement." Manuscript of a chapter submitted to the AASA, NSBA and NEA-Classroom Teachers Division for incorporation into a book titled *The Evaluation of Teaching*. (Personal communication from Dr. Flanders.) 1964b.

32. ______. "Interaction Models of Critical Teaching Behaviors," *An Analysis and Projection of Research in Teacher Education*. Cooperative Research Project No. F-015. Frederick R. Cyphert and Ernest Spaights, editors. Columbus, Ohio: The Ohio State University Research Foundation, 1964.

33. FURST, NORMA. "The Effects of Training in Interaction Analysis on the Behavior of Student Teachers in Secondary Schools." Paper presented at the thirty-eighth annual meeting of the American Educational Research Association, Chicago, February 1965.

34. GAGE, N. L. "A Method for 'Improving' Teacher Behavior," *Journal of Teacher Education*, 14:261-6, September 1963.

35. GARRETT, HENRY E. *Statistics in Psychology and Education*. Fifth edition. New York: David McKay Company, Inc., 1965.

36. GUILFORD, J. P., and P. R. MERRIFIELD. *The Structure of Intellect Model: Its Uses and Implications*. Report No. 24. Los Angeles: The Psychological Laboratory of the University of Southern California, 1960.

37. HARE, A. PAUL, EDGAR F. BORGATTA, and ROBERT F. BALES. *Small Groups: Studies in Social Interaction*. New York: Alfred A. Knopf, Inc., 1955.

38. HOUGH, JOHN B., and EDMUND J. AMIDON. "The Relationship of Personality Structure and Training in Interaction Analysis to Attitude Change During Student Teaching." Paper presented at the thirty-eighth annual meeting of the American Educational Research Association, Chicago, February 1965.

39. HUGHES, MARIE M., *et al*. "Teaching is Interaction," *Elementary School Journal*, pp. 457-64, May 1958.

40. ______, *et al*. *Assessment of the Quality of Teaching in Elementary Schools: A Research Report*. Salt Lake City: University of Utah, 1959.

41. HURD, ARCHER WILLIS. "Progress Report on the Development of Teaching Units in High School Physics," *The North Central Association Quarterly*, September 1930.

42. ______. *Cooperative Experimentation in Materials and Methods in Secondary School Physics*. New York: Columbia University, Teachers College, Bureau of Publications, 1933.

43. ______. "Experimental Efforts to Improve the Teaching of Physics," *Science Education 18, No. 1*, February 1934.

44. JOHNSON, PHILIP G. (ed.). *Supervisory Needs and Recommended Actions for Improving Science and Mathematics Instruction*. A report of the 1962 Cornell-Shell Supervisors Conference. Ithaca, New York: Cornell University, Science Education Division, 1962.

45. ______. "Evaluating Science Teachers," *The Science Teacher*, Vol. 30, No. 7, 1963a.

46. ______. "Emerging Curriculum Studies in Elementary and Junior High School Science," *Supervision for Quality Education in Science*. Report of a conference arranged by the Specialists For Secondary Science, June 25-29, 1962. Washington, D. C.: U. S. Department of Health, Education, and Welfare, Office of Education, 1963.

47. KLEINMAN, GLADYS S. "General Science Teachers' Questions, Pupil and Teacher Behaviors, and Pupils' Understandings of Science." Paper presented at the thirty-eighth annual meeting of the National Association for Research in Science Teaching, Chicago, February 14, 1965.

48. MALLINSON, GEORGE G. *Final Report on Science Motivation Project I, Cooperative Research Project 503: An Analysis of the Factors Related to the Motivation and Achievement of Students in Science Courses in the Junior and Senior High School*. Washington, D. C.: U. S. Office of Education, Cooperative Research Program.

49. MATTHEWS, CHARLES C. "The Classroom Verbal Behavior of the Secondary School Science Student Teacher and the Cooperating Classroom Teacher." Unpublished doctoral dissertation, Cornell University, Ithaca, New York, 1965.

50. MEDLEY, DONALD M. "Experiences With the OScAR Technique," *Journal of Teacher Education*, 14:267-73, September 1963.

51. ______, and HAROLD E. MITZEL. "A Technique for Measuring Classroom Behavior," *Journal of Educational Psychology*, 49:86-92, 1958.

52. ______, and ______. "Some Behavioral Correlates of Teacher Effectiveness," *Journal of Educational Psychology*, 50:239-46, 1959.

53. ______, and ______. "Measuring Classroom Behavior by Systematic Observation," in *Handbook of Research on Teaching*, N. L. Gage, editor. American Educational Research Association. Chicago: Rand McNally and Company, 1963.

54. MORRISON, VIRGINIA B. "The Relationship fo Student Teacher Performance and Pupil Performance to Supervisory and Pupil Merit Ratings." Unpublished doctoral dissertation, University of Michigan, Ann Arbor, 1961.

55. NATIONAL SCIENCE FOUNDATION. Proceedings of the Ad Hoc Advisory Conference on Secondary School Physics, 1963.

56. ORT, VIRGIL K. "A Study of Some Techniques Used for Predicting the Success of Teachers," _Journal of Teacher Education_, 15:67, 1964.

57. PARAKH, JAL S. "A Study of Teacher-Pupil Interaction in High School Biology Classes." Unpublished doctoral dissertation, Cornell University, Ithaca, New York, 1965.

58. REED, HORACE B. "Teacher Variables of Warmth, Demand and Utilization of Intrinsic Motivation Related to Pupils' Science Interests," _Journal of Experimental Education_, 29:205-29, March 1961.

59. RIPPLE, RICHARD E., and VERNE N. ROCKCASTLE (eds.). _Piaget Rediscovered_. A Report of the Conference on Cognitive Studies and Curriculum Development. Ithaca, New York: Cornell University, School of Education, 1964.

60. RYANS, DAVID G. _Characteristics of Teachers_. Washington, D. C.: American Council on Education, 1960.

61. ______. "Assessment of Teacher Behavior and Instruction," _Review of Educational Research_ _33_, _No_. _4_, October 1963.

62. SEARLE, S. R. "Correlation Matrix (CORMA)." Unpublished documentation of CORMA Program for CDC 1604 computer, Cornell Computing Center. Programmers: S. R. Searle and P. L. Primer. 1964a.

63. ______. "Multiple Regression Analyses (MUREG)." Unpublished documentation of MUREG Program for CDC 1604 computer, Cornell Computing Center. Programmer: P. L. Primer. 1964b.

64. ______. "Rank Correlation Analyses (RANKO)." Unpublished documentation of RANKO Program for CDC 1604 computer, Cornell Computing Center. Programmer: P. L. Primer. 1964c.

65. SIEGEL, SIDNEY. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill, 1956.

66. SMITH, B. OTHANEL. *A Study of the Logic of Teaching: A Report on the First Phase of a Five-Year Research Project*. Washington, D. C.: U. S. Office of Education, 1959.

67. ______. "Toward A Theory of Teaching," in *Theory and Research in Teaching*, Arno A. Bellack, editor. New York: Columbia University, Teachers College, Bureau of Publications, 1963a.

68. ______. "A Conceptual Analysis of Instructional Behavior," *Journal of Teacher Education*, 14:294-98, September 1963b.

69. THELEN, HERBERT A. "Insights for Teaching From A Theory of Interaction," in *The Nature of Teaching*, Louise M. Berman, editor. Milwaukee, Wisconsin: University of Wisconsin, 1963.

70. THOMAS, DOROTHY S., *et al*. "Some New Techniques for Studying Social Behavior," *Child Development Monographs*, 1929.

71. TRENT, JOHN HENRY. "The Attainment of the Concept 'Understanding Science' Using Contrasting Physics Courses." Paper presented at the 1965 convention of the National Association for Research in Science Teaching, Chicago, 1965.

72. VAN HOOFT, GORDON E. (ed.). Physics. An Experimental Syllabus. 1964 edition. The University of the State of New York. Albany, New York: State Education Department, Bureau of Secondary Curriculum Development, 1964.

73. WATSON, FLETCHER G. "Research on Teaching Science," Chapter 20 in *Handbook of Research on Teaching*, N. L. Gage, editor. American Educational Research Association, 1963.

74. WIGHTMAN, LAWRENCE E. "Interaction Analysis Program." Unpublished documentation of program for Flanders' system of interaction analysis, for CDC 1604 computer, revised. Programmer: L. Wightman. Cornell Computing Center, Learning Structures Project, Cornell University, 1965.

75. WISPE, L. G. "Evaluating Section Teaching Methods in the Introductory Course," *Journal of Educational Research*, 45:161-186, 1951.

76. WITHALL, JOHN. "Development of a Technique for the Measurement of Socio-Emotional Climate in Classrooms," *Journal of Experimental Education*, 17:347-61, 1949.

77. ______. "The Development of a Climate Index," *Journal of Educational Research*, 45:93-99, 1951.

78. ______, and W. W. LEWIS. "Social Interaction in the Classroom," Chapter 13 of *Handbook of Research on Teaching*, N. L. Gage, editor. Chicago: Rand McNally and Company, 1963.

79. WRIGHT, E. MURIEL J., and VIRGINIA H. PROCTOR. "Systematic Observation of Verbal Interaction As A Method of Comparing Mathematics Lessons." U. S. Office of Education Cooperative Research Project No. 816. St. Louis, Missouri: Washington University, 1961. (Mimeographed.)

80. ZAHN, RICHARD D. "The Effect Upon Student Teachers' Attitudes of Training in Interaction Analysis and the Attitudes of Cooperating Teachers." Paper presented at the thirty-eighth annual meeting of the American Educational Research Association, Chicago, February 1965.

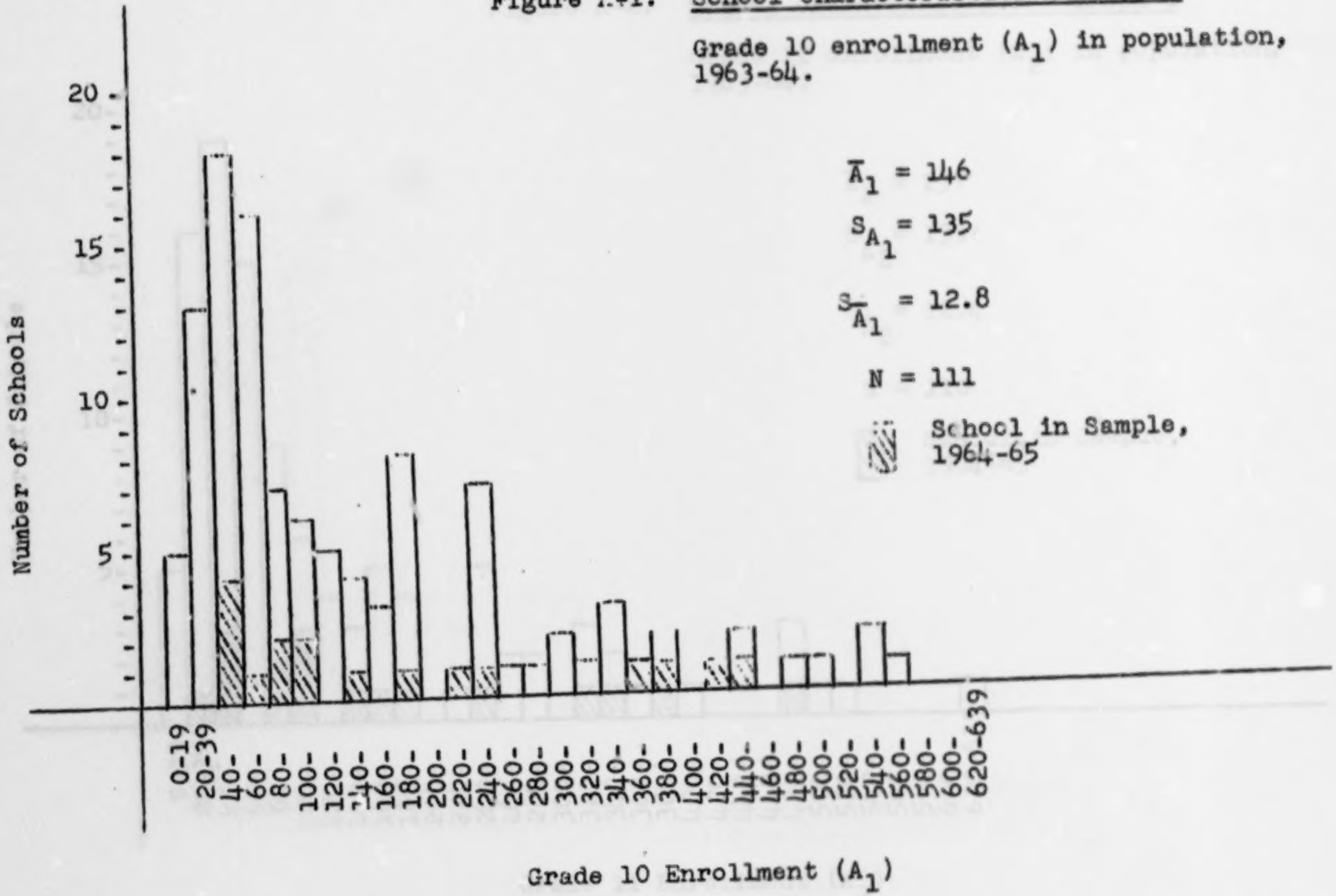
APPENDIX A

FREQUENCY DISTRIBUTIONS OF SAMPLE AND POPULATION DATA ON TWELVE SELECTED CHARACTERISTICS OF TEACHERS AND SCHOOLS

Figures A-1 through A-12

Figure A-1. School Characteristics - Graph 1

Grade 10 enrollment (A_1) in population, 1963-64.



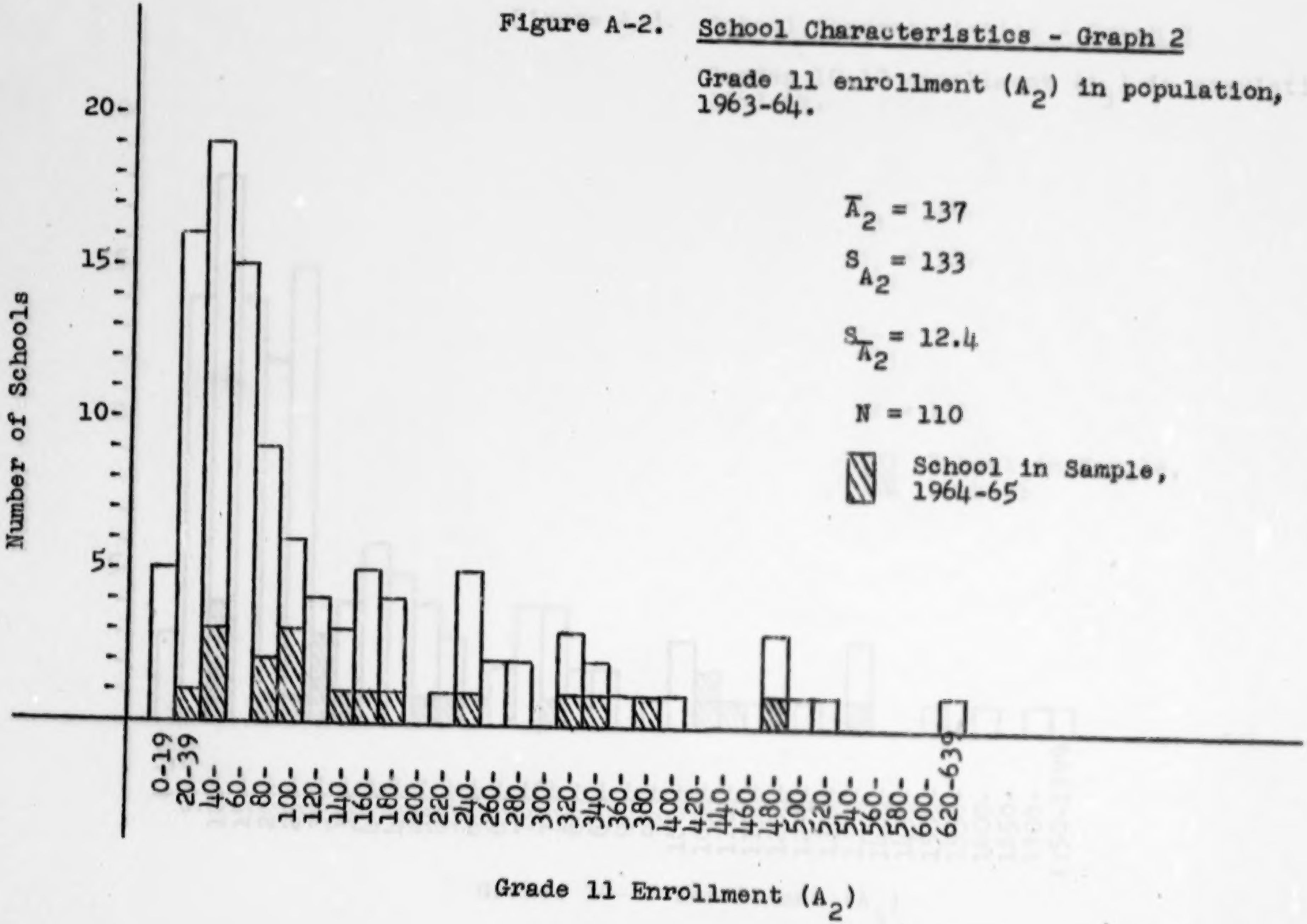


Figure A-2. School Characteristics - Graph 2

Grade 11 enrollment (A_2) in population, 1963-64.

$\bar{A}_2 = 137$

$S_{A_2} = 133$

$S_{\bar{A}_2} = 12.4$

$N = 110$

Figure A-3. School Characteristics - Graph 3

Grades 10-12 enrollment (A_3) in population, 1963-64.

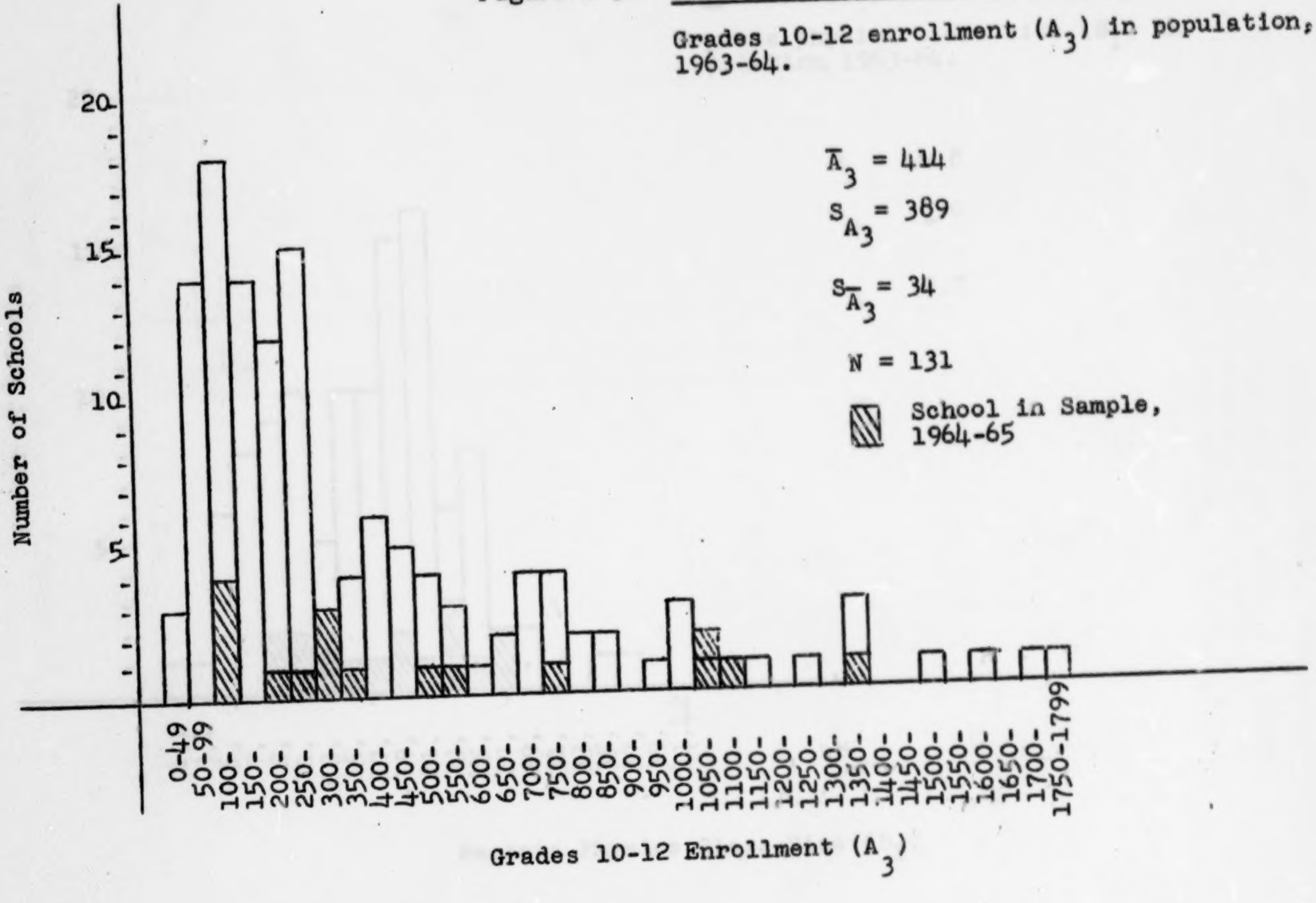


Figure A-4. School Characteristics - Graph 4

Regents physics class size (B_1) in population, 1963-64.

$\overline{B}_1 = 16.8$

$S_{B_1} = 7.0$

$S_{\overline{B}_1} = 0.7$

N = 111

▨ School in Sample, 1964-65

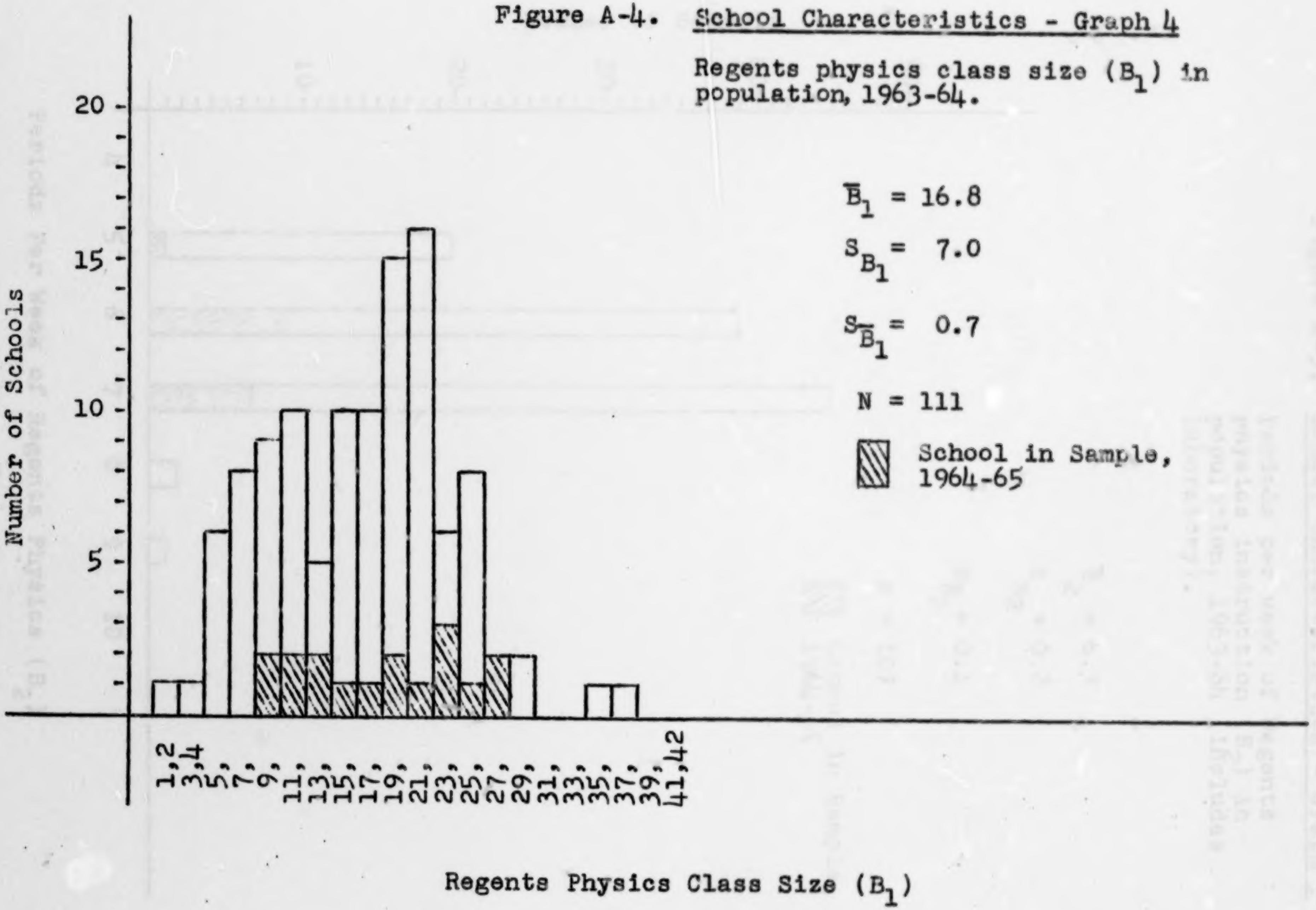


Figure A-5. School Characteristics - Graph 5

Periods per week of Regents physics instruction (B_2) in population, 1963-64 (includes laboratory).

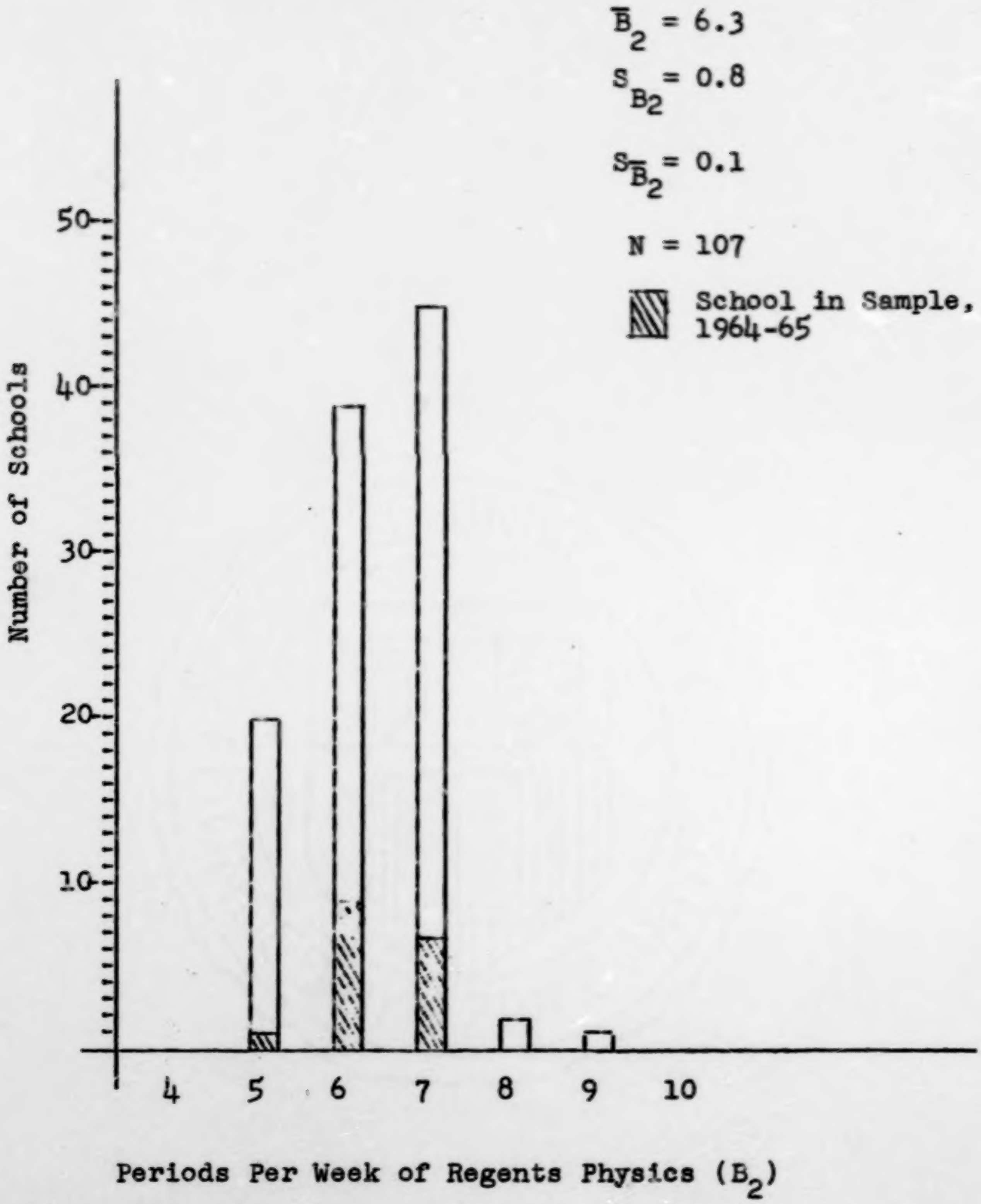


Figure A-6. School Characteristics - Graph 6

Percent physics enrollment is of grades 10-12 enrollment (B_3) in population, 1963-64.

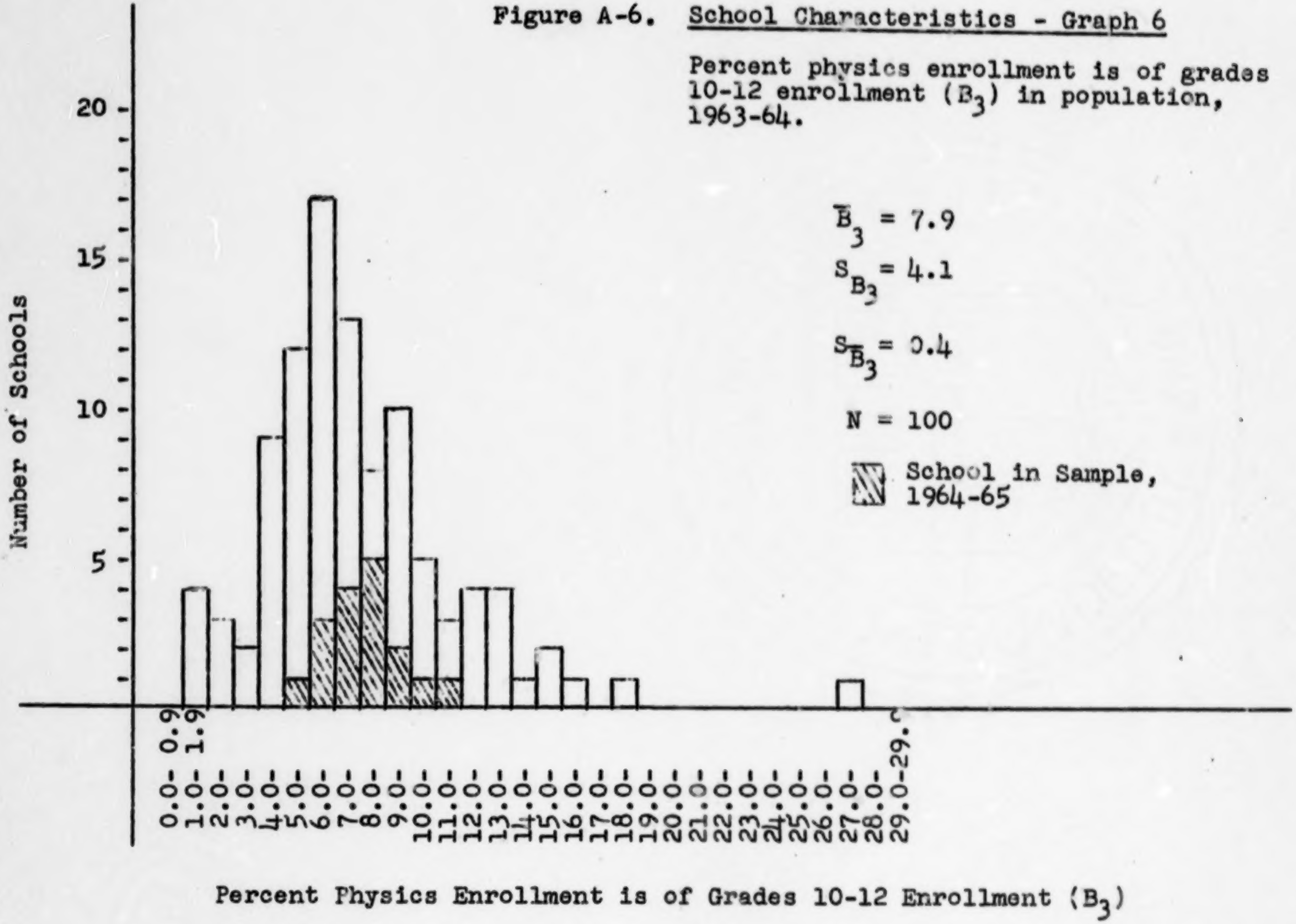


Figure A-7. School Characteristics - Graph 7

Physics teachers' years of experience in secondary science teaching (C_1) in population, 1963-64.

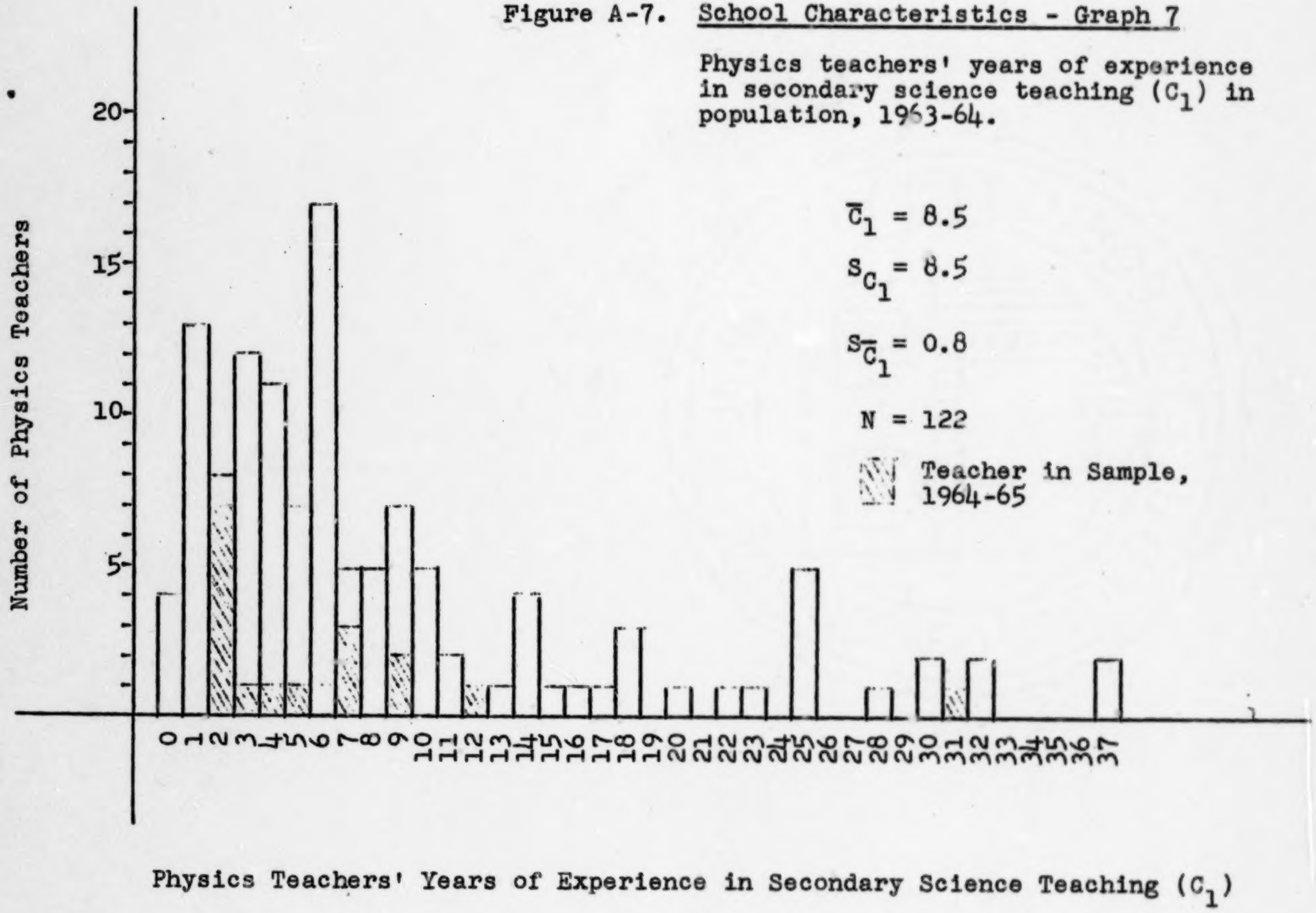
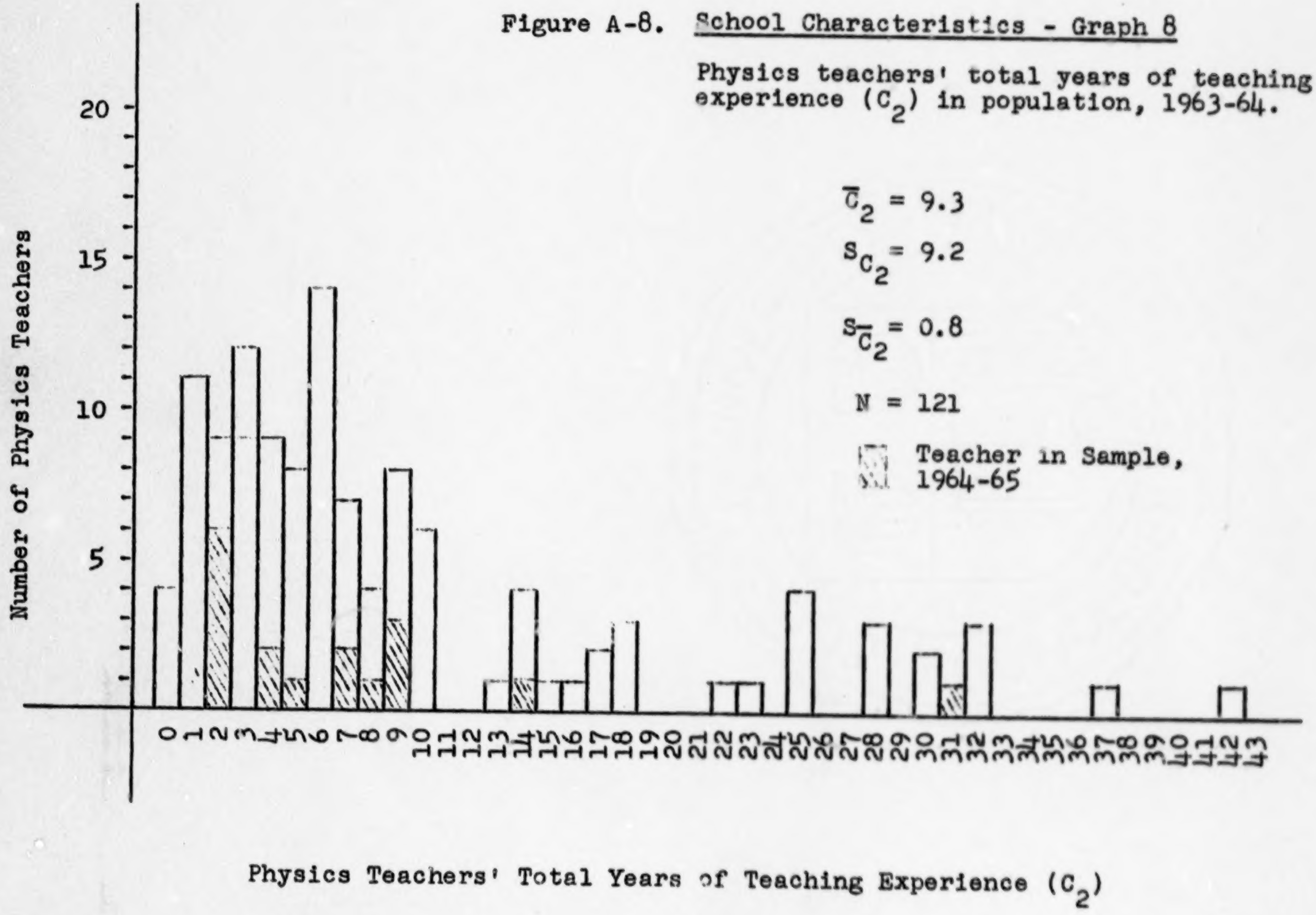


Figure A-8. School Characteristics - Graph 8

Physics teachers' total years of teaching experience (C_2) in population, 1963-64.



$\overline{C}_2 = 9.3$

$S_{C_2} = 9.2$

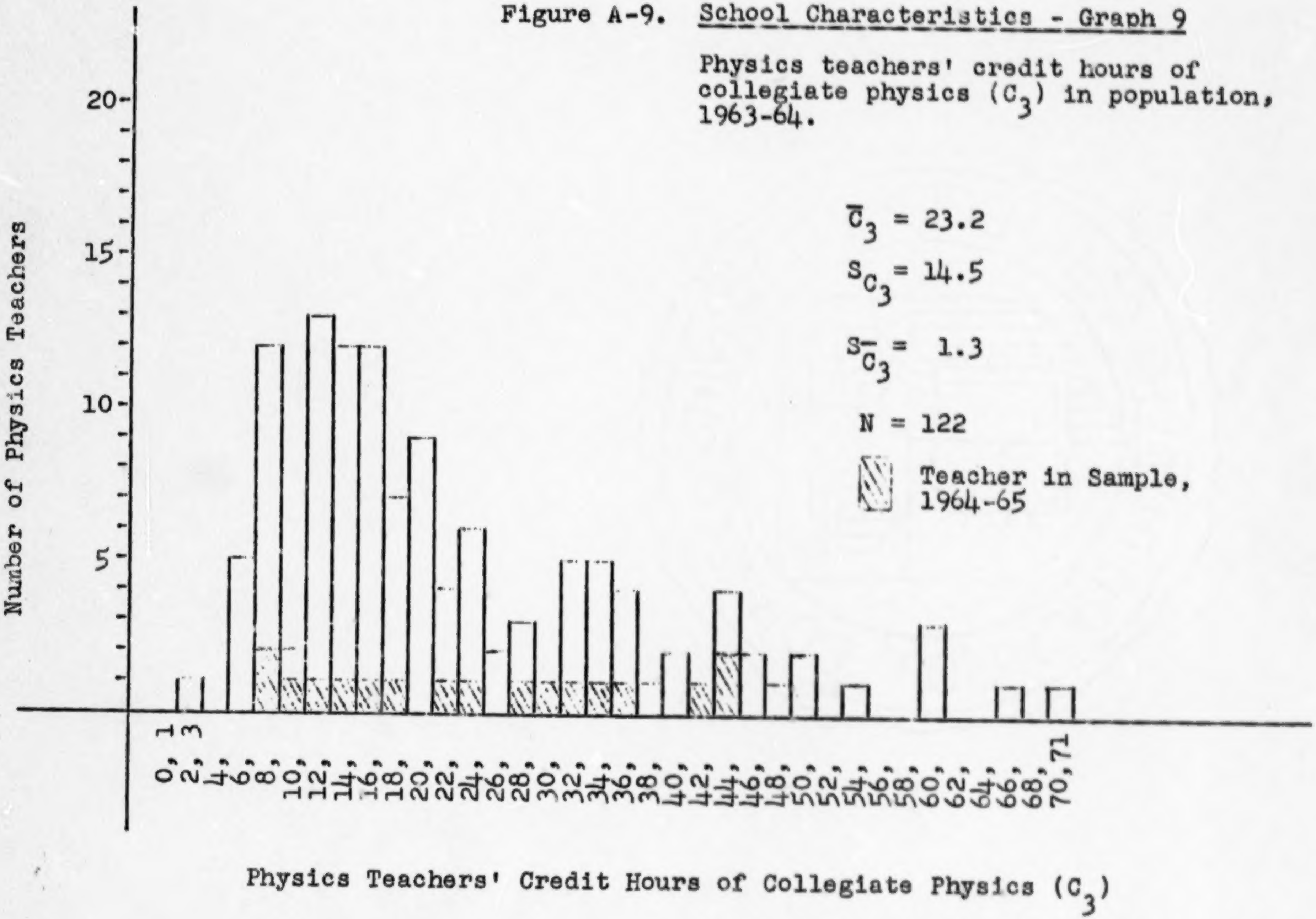
$S_{\overline{C}_2} = 0.8$

$N = 121$

▨ Teacher in Sample, 1964-65

Figure A-9. School Characteristics - Graph 9

Physics teachers' credit hours of collegiate physics (C_3) in population, 1963-64.



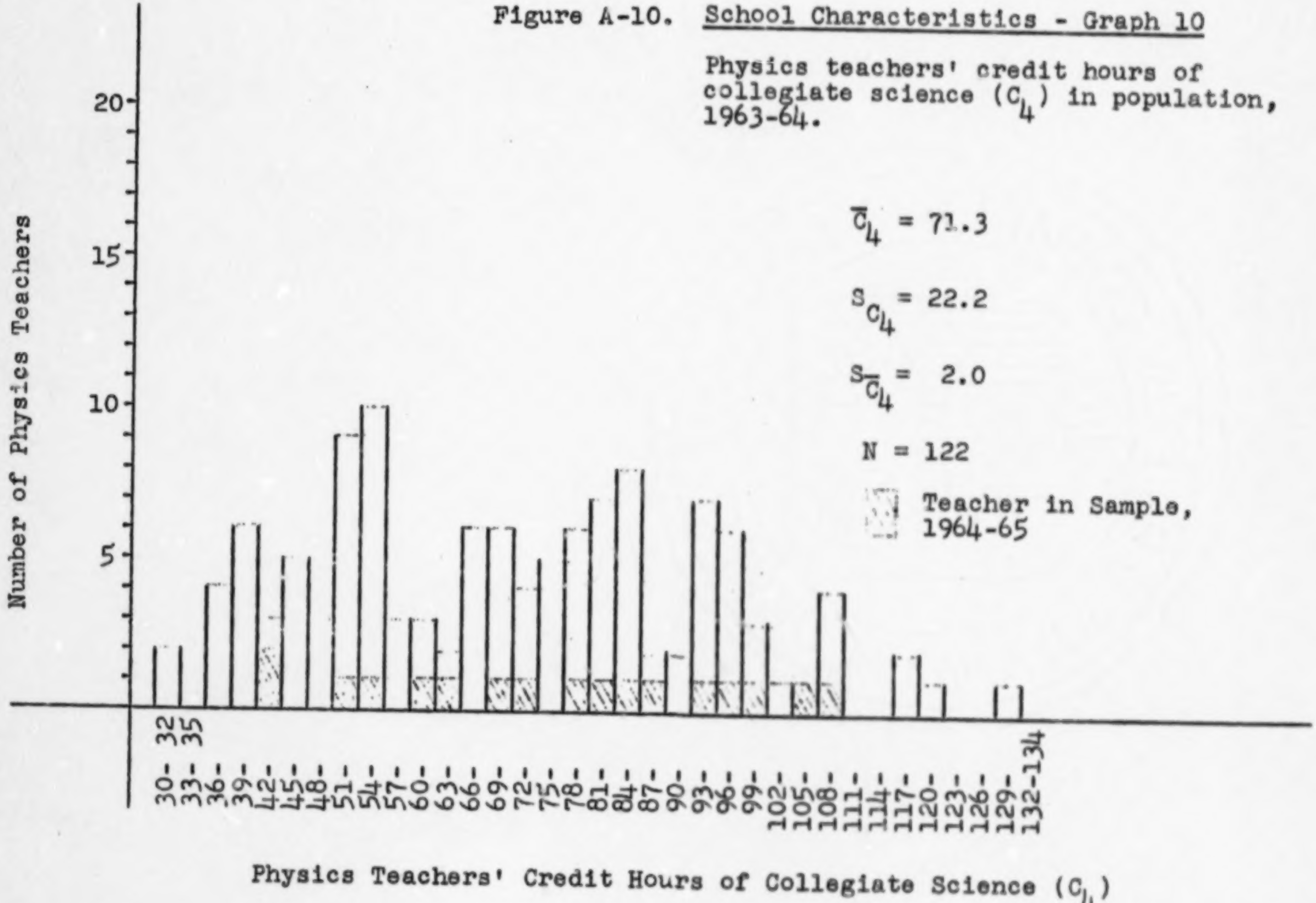
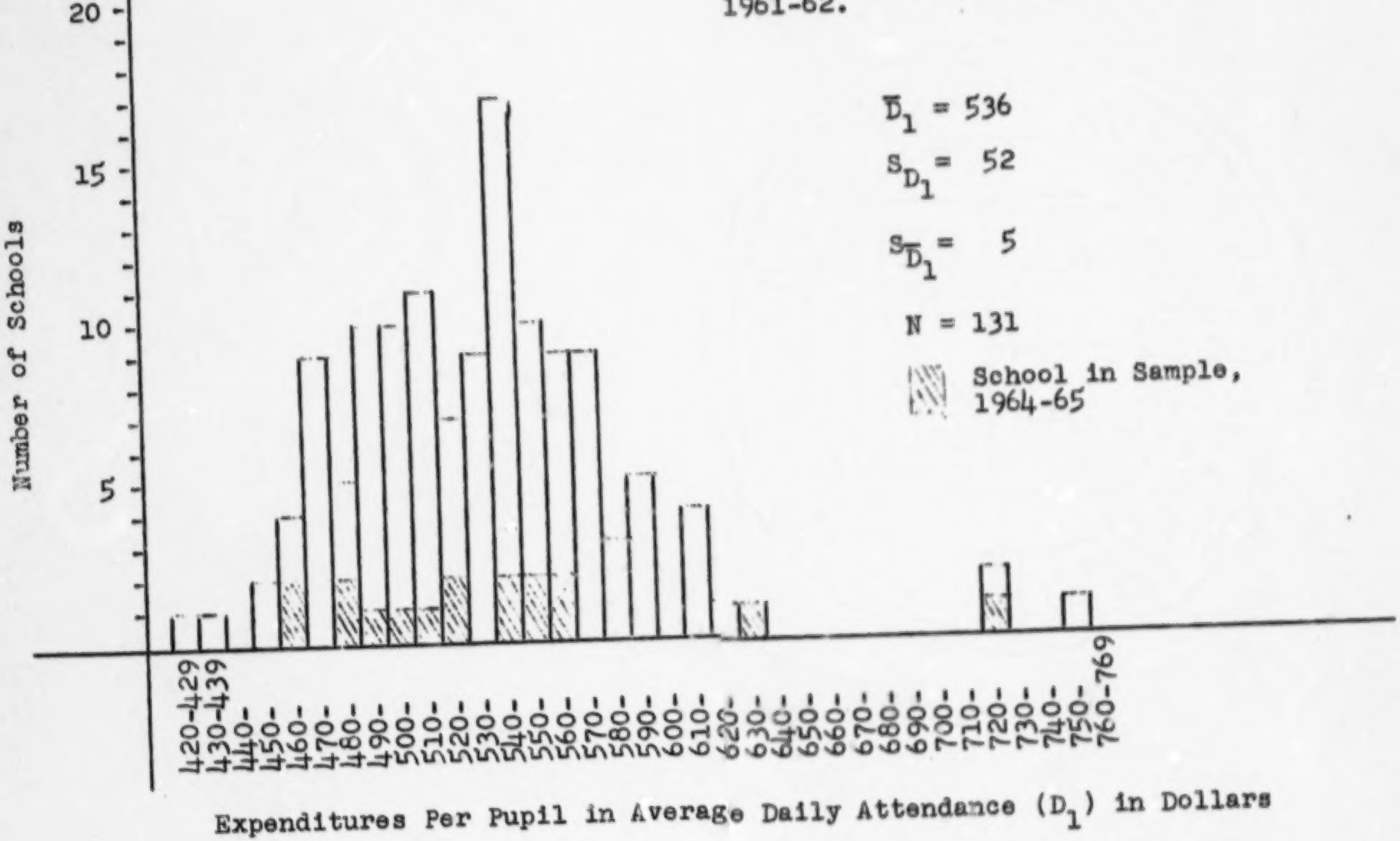


Figure A-10. School Characteristics - Graph 10

Physics teachers' credit hours of collegiate science (C_4) in population, 1963-64.

Figure A-11. School Characteristics - Graph 11

Districts' expenditures per pupil in average daily attendance (D_1) in population, 1961-62.



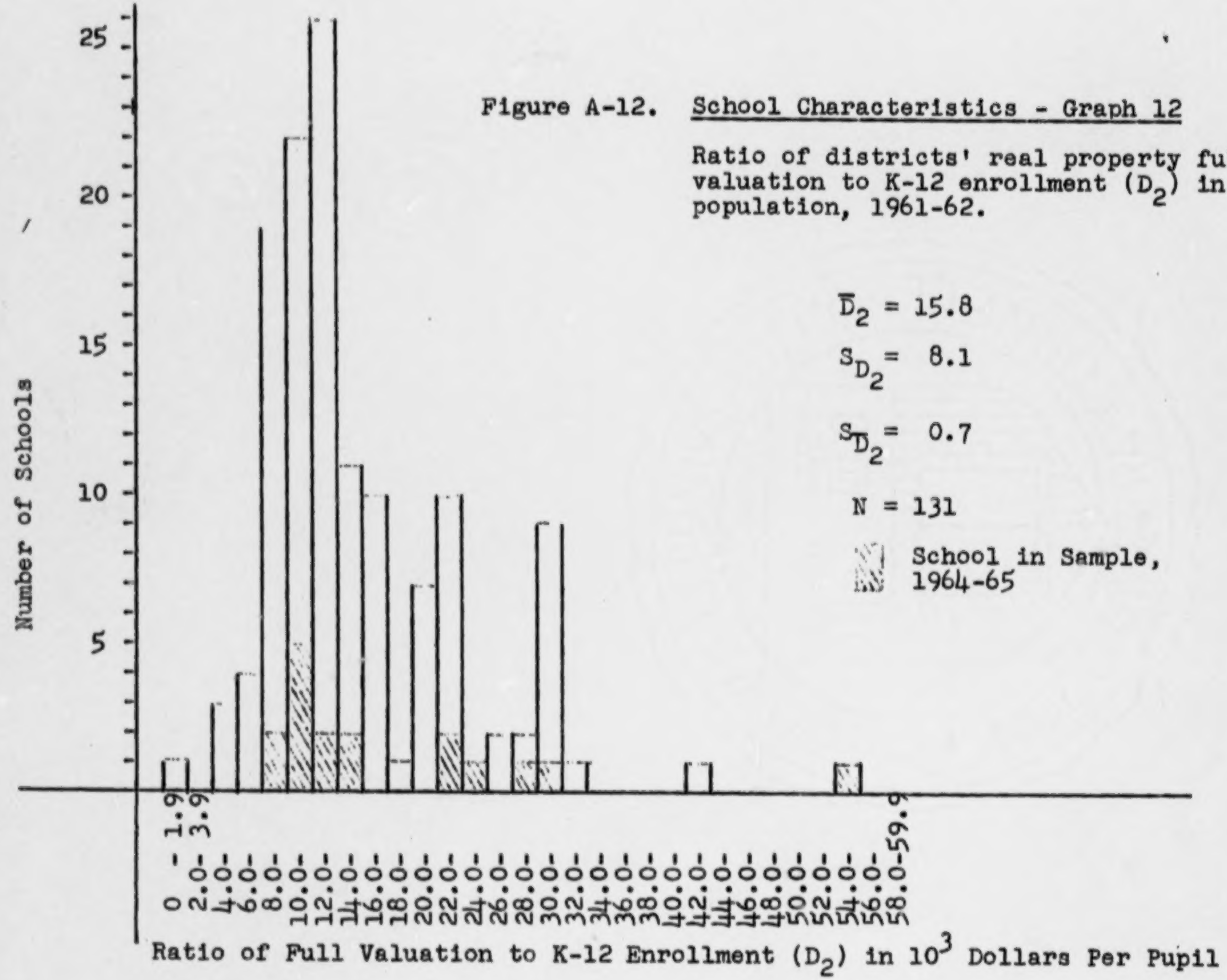


Figure A-12. School Characteristics - Graph 12

Ratio of districts' real property full valuation to K-12 enrollment (D_2) in population, 1961-62.

APPENDIX B

LETTERS AND MEMORANDA

Form B-1	Solicitation Letter to Principals
Form B-2	Solicitation Letter to Teachers
Form B-3	Memorandum to Physics Teachers, Science Department Heads, and Principals
Form B-4	Memorandum Accompanying "Summary of Data on School Characteristics" Report
Form B-5	Memorandum Accompanying "Summary of Data on School Characteristics"--Special Form to Teachers Only

Form B-1

10 Stone Hall
Division of Science Education
Cornell University
Ithaca, New York
(Date)

Principal's name
School address
City, New York

Dear Mr. (principal's name):

I am a doctoral candidate in Science Education at Cornell University developing a thesis on physics teaching. I would like very much to have your physics teacher, (name of teacher), participate in the study.

A number of changes involving secondary school physics have been made in the last ten years. Reliable information concerning teacher and student behavior in the physics classroom is relatively scarce. There is a need to study teaching in the schools where it is regularly done.

My major purpose is to attack the question, "What is the nature of physics teaching?" My approach will be to observe physics teaching in selected schools in a central New York State region. A modification of the Flanders technique of interaction analysis will be used. No teacher will be judged and there will be no need to change normal operations.

I wish to observe one Regents physics class for four consecutive days in the fall and for three consecutive days in the spring. I hope to perfect an instrument which can describe physics teaching with consistent results.

I would like to meet you and Mr. (teacher's name), preferably together, to discuss my proposed research and to ask for your cooperation. I have written him about this. May I please have an appointment during the week of October 5? To save you the trouble of writing, I will phone your office this coming Friday afternoon.

Thank you in advance for your consideration.

Very truly yours,

Ray M. Snider
Principal Investigator

cc: (Teacher's name)

This study is under the guidance of a graduate committee of which I am chairman. Your cooperation will be appreciated.

Philip G. Johnson
Chairman, Division of Science Education

Date Form B-2

Teacher's name
School address
City, New York

Dear (teacher's name):

I have written your principal, Mr. (principal's name), concerning my research on physics teaching. I would like very much to observe one of your Regents physics classes as indicated in the attached copy.

Recent research points to the need to study teaching in the classroom if progress is to be made in the development of a better understanding of its nature. It is important to point out that my major concern is to take a beginning step in this direction by developing a means of describing physics teachers like yourself.

I will be very grateful for an opportunity to meet with you and Mr. (principal's name) in this regard. Will you please contact him so that an appointment may be arranged?

Thank you very much.

Very truly yours,

Ray M. Snider
Principal Investigator

MEMORANDUM

Form B-3

To: Physics Teachers, Science Department Heads, and Principals
From: Ray M. Snider
Re: Research Project No. S-2103 entitled "Teacher-Student Verbal Interaction in the High School Physics Classroom"

The following sketches pertinent aspects of the study. COMPLETE ANONYMITY AND THE HIGHEST DEGREE OF PROFESSIONAL ETHICS WILL PREVAIL.

A. General Design.

The study is focused on high school physics teaching. A major objective is to develop a method of analyzing classroom discourse. To carry out the research, it is necessary to go into high school classrooms and observe physics teaching in a purely objective fashion. No teacher or school will be rated in any way. Some testing is necessary to analyze student learnings and attitudes.

B. Procedures.

1. Observations: A sample of 17 Regents physics teachers and their students in 17 schools of a 14-county region (surrounding Ithaca) will be studied. The sample will be constructed so as to reflect the region on characteristics such as enrollment, cost per pupil, teacher's experience, etc.

(a) One class of each participating teacher will be observed for four consecutive days in the fall and for three consecutive days in the spring. All observations, analyses, and reporting will be carried out by Mr. Snider.

(b) During observation, Mr. Snider will record notations while tape recording the discourse. A portable battery-powered recorder will be used. No attempt will be made to deceive students; however the recorder will be in a briefcase and students will not be distracted by gadgetry and spinning reels.

(c) Teachers are asked not to give a test longer than 10 minutes, nor take a field trip, nor show a motion picture longer than 10 minutes during the observation. Otherwise, they are urged to do just as they ordinarily would if the observer were not present. It is important that no special lessons be arranged for the observer's visits.

(d) It is hoped to complete each cycle of visits in 10 weeks. A schedule will be arranged as soon as the sample is complete.

2. Testing: All test materials will be provided by

Mr. Snider. They will be mailed to the teachers with return postage. Teachers are asked to administer the tests.
(a) The "Cooperative Physics Test, Form Z" will be administered during the week of October 19 and again during the week of May 17.
(b) The "Test On Understanding Science" will be administered during the week of October 19 or week of October 26 and again during the week of May 17 or week of May 24. Both tests require 40 minutes.
(c) The "Minnesota Student Attitude Inventory" will be mailed to students individually as a questionnaire and with return postage. This will be done in early February, 1965.

3. Other: The following is also requested:
(a) Access to student records so that I.Q. and previous achievement data may be collected.
(b) Regents test scores in physics in June, 1965.

(c) The current year's science enrollment figures ("brown card" data).

C. In return for the teacher's and school's cooperation and assistance as outlined above, Mr. Snider will render the following services:

1. Summary reports of the study will be produced and delivered by August 15, 1965 (two copies per school--more if requested).

2. As soon as tabulations are complete, a statistical survey of characteristics of all schools in the 14-county region will be specially produced for the participating schools. (Three copies per school--mimeographed and delivered by mid-November, 1964. Additional copies if requested.)

3. Teachers will receive pre, post and gain scores on CPT and TOUS by July 1, 1965.

4. Other requests which do not interfere with the experimental design will be considered.

D. Snider vita:
1. Home address: 54-H Hasbrouck Apts., Ithaca, N. Y.
Phone: AR3-5941
Office: 10 Stone Hall, Cornell University, Ithaca, New York.
Phone: AR5-5410; AR5-2379
(collect calls will be graciously accepted)

MEMORANDUM (Continued) Form B-3

2. Ph.D. candidate in Science Education, Cornell University. Major: Science Education; Minor: Physics; Minor: Math. Age 34. Married - 2 children.

3. Experience: 5 years physics and mathematics teaching at Roosevelt High School, Hyde Park, New York (Gilbert B. Pearson, Principal); 1 year college physics teaching at Oklahoma State University; teaching of physics laboratory in Shell Merit Fellowship Program at Cornell (2 summers); research assistant in Cornell's "Teacher Research Experience Program," 1962-1963. BS, Albany State ('56); MS, RPI ('60).

4. Plans: Science teacher training and physics teaching in New York State University system commencing fall, 1965.

E. The study outlined is part of Mr. Snider's doctoral dissertation. It is hoped to receive support from the U. S. Office of Education. Approval and support of Cornell's Division of Science Education, Dr. Philip G. Johnson, Chairman, is already granted. Verbal approval from the State Education Department has been received as well. Your assistance and cooperation will be greatly appreciated.

Ray M. Snider
Principal Investigator

MEMORANDUM Form B-4

To: Participating teachers and administrators of schools taking part in research project

From: Ray Snider

Re: Summary of data on school characteristics

Date: January 1965

In carrying out my current study, in which you are participating, I found it necessary to collect a quantity of data on my so-called "population" of 131 public high schools in central New York State. This data will serve the purpose of showing that my sample of 17 participating schools is as good, statistically, as a random sample of schools. My use of the data essentially ends at that point.

In the early stages of my work, it occurred to me that a summary of such data might be one of several things that I could produce for participating schools in return for their cooperation in the study. With this in mind, and, not knowing to what extent such a summary will be useful to you, I have nevertheless produced the accompanying report. Please don't hesitate to communicate with me if you have questions concerning the latter's content.

As you know, the phases of fall observation and fall testing have been completed. Spring observation (3 consecutive days) and testing remain. A summary report on the entire study will be delivered by mid-August, 1965.

Again, my sincere thanks to you for your willingness to participate in my research study. Although somewhat belated, allow me to extend my best wishes for a fine New Year.

MEMORANDUM Form B-5

To: Participating teachers

From: Ray Snider

Re: Summary of data on school characteristics; and school calendar

A. Enclosed you will find the following:

(1) Three copies of "Summary of Data on Selected Characteristics of Public High Schools Comprising a Fourteen-County Central New York State Region

(2) Three copies of an associated memorandum to teachers and administrators

(3) One stamped and addressed envelope

Item (2) provides an indication of the nature of (1). Will you kindly give two copies of the "Summary . . ." and two copies of the associated memorandum to your high school administrator(s)? The third copy of each is for your files. I will gladly forward additional copies if such are requested.

B. I am currently preparing to arrange my schedule of spring visits. These will be carried out in the same manner as the fall observations--however, they will run for three consecutive days instead of four. To facilitate such an arrangement it is necessary for me to obtain a copy of your school calendar (for the spring semester, at least). Will you please send me one copy as soon as possible? The stamped and addressed envelope is for your convenience. Of course I will solicit your approval of a tentative schedule before final plans are made.

Thanks very much for your help.

APPENDIX C

QUESTIONNAIRES USED

C-1 Teacher Questionnaire

C-2 Student Questionnaire

C-3 Modifications of Michigan Student Questionnaire Items

C-4 Memorandum on Student Questionnaire

TEACHER QUESTIONNAIRE

C-1

Code ____
June, 1965

Dear Participating Teacher:

Your responses to the following items will be very helpful in guiding future research efforts involving classroom observation. In each case, consider the six or more observations as a whole. It is very important that you be completely frank and sincere in your responses. No one will see individual returns but me. Please answer every item by placing an X on the appropriate letter. Additional remarks are invited. Thank you very much.

1. After my first visit, how distracted were you with my presence in the classroom?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

2. After my first visit, how distracted was your class with my presence in the classroom?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

3. To what extent, when I was there, and after the first visit, did your class refrain from participating in class discussions because of my presence in the classroom?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

4. To what extent, when I was there, and after the first visit, do you think you taught differently from the way you usually teach?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

5. To what extent, after the first visit, were you hindered: by an awareness of the fact that your classroom discourse was being recorded on tape?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

by not being able to give a test (or show a film) of duration greater than ten or fifteen minutes?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

6. To what extent did your class, as a whole, appear to make a serious effort in answering items on CPT and TOUS: in the fall?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

in the spring?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

7. Now that it's over, considering your year's participation in my project, to what extent were you inconvenienced (on the whole)?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

* *

Please feel free to make any additional comments that you care to. You may use the other side for more space if you wish. (Note: The teacher questionnaire in the original format was reproduced on one side of one page.)

A. Aspects of the operation that were inconvenient:

B. Other:

C-2 Spring, 1965

Dear High School Physics Student:

This questionnaire is used as part of my research on high school physics teaching. It is part of the same project that is connected with my visits to your physics class in the fall and, in most cases, in the spring. I am very grateful to you and to your physics teacher for your participation in my research study. You may be assured that your responses will be completely confidential and will not, in any way, jeopardize you, your physics teacher or your school. I will appreciate very much your prompt and serious attention to it. Thank you.

Ray M. Snider
Ph.D. Candidate in Science Education
Cornell University

* *

STUDENT QUESTIONNAIRE

School _______ Male ( ) Female ( )
(Respondent please check.)

This is not a test because there are no wrong answers. The answer to each question is A MATTER OF OPINION, and your true opinion, whatever it is, IS THE RIGHT ANSWER. You will be asked a lot of questions about how much you like your physics class, the teacher, and the work you are doing there. All the items refer to THAT ONE CLASS AND THAT PARTICULAR TEACHER. No one in your school will see your answers. By giving frank, true answers to show exactly how you feel, you can help us understand the opinions of students.

DIRECTIONS:

1. Complete the questionnaire outside of class (preferably at home) as soon as you can and mail it using the stamped and addressed envelope provided. There is no need to record your name or to give a return address on the envelope. Your class will be identified by code letters already written in.
2. Do not skip any items. Work carefully, but quickly. Don't spend too much time deciding how to answer each item--mark the answer that comes to your mind first. Don't confer with a friend, just give your own opinion.
3. For each item, mark the abbreviation that is most nearly like your opinion. You may use pencil or pen. Make one choice of the four alternatives. Be sure to check either "male" or "female" above.

STUDENT QUESTIONNAIRE (Continued)

EXAMPLES:

0. I think we should have school on Saturday.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

You have four alternatives to choose from. If you STRONGLY DISAGREE with the statement, draw a circle around SD immediately following the statement (as above).

00. Girls talk more than boys do.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

If you aren't really certain about this, but you are inclined to AGREE, you should draw a circle around the A immediately following the statement. However, if you STRONGLY AGREE, you should encircle SA instead. On the other hand, if you aren't really certain about it, but are inclined to DISAGREE, then you should encircle D.

* *

QUESTIONS:

1. I get along well with this teacher.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

2. This teacher has lots of fun with us.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

3. This teacher is good at thinking things through.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

4. This teacher helps to settle quarrels fairly.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

5. I'm usually afraid that I will fail my tests.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

6. This teacher lets some students get by without working very hard.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

7. This teacher praises us for good work.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

8. This teacher gives all of us opportunities for sharing in enjoyable tasks.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

9. I think this teacher picks on some students unfairly.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

STUDENT QUESTIONNAIRE (Continued)   C-2

10. We don't pay much attention in this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

11. This teacher lets us discuss things in class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

12. We behave well in this class even when the teacher is out of the room.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

13. This teacher will always listen to both sides of an argument.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

14. This teacher is quick to see what mixes you up in your schoolwork.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

15. This teacher makes sure WE understand our work.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

16. This teacher is always fair with each student.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

17. I worry about getting good grades.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

18. This teacher always asks the OTHER students the EASY questions.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

19. What we learn in this class makes me want to learn new things.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

20. This teacher is one of the best I have ever had.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

21. I get pretty bored in this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

22. This teacher sometimes punishes the whole class for something one person did.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

23. I never worry about taking tests.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

24. This teacher certainly knows how to teach.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

STUDENT QUESTIONNAIRE (Continued) C-2

25. This teacher really understands students my age.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

26. This teacher knows a lot.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

27. I find it easy to talk with this teacher.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

28. Our teacher makes everything seem interesting and important.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

29. I worry about getting my work done.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

30. I'm afraid to raise my hand to talk in this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

31. Our teacher never gives us extra assignments as punishment.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

32. This teacher makes sure not to hurt your feelings.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

33. This teacher often "bawls you out" in front of the class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

34. We often complain just to get out of work.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

35. I really like this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

36. The students in this class behave well without being told.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

37. This teacher is not easily upset.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

38. I like to be called on in this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

39. This teacher keeps order with a fair and firm hand.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

40. This teacher makes it fun to study things.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

41. This teacher never asks our opinion in planning work to be done.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

42. This teacher doesn't listen to what SOME students have to say.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

43. This teacher never gets angry and shouts at us.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

44. Our teacher helps us when we have problems with our work.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

45. This teacher has some special favorites or "teacher's pets".
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

46. This teacher makes me nervous.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

47. This teacher helps us get the most out of each day.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

48. This teacher likes teen-agers.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

49. I wish I could have this teacher next year.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

50. This teacher likes to hear students' ideas.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

51. This teacher makes sure no students get left out of things.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

52. Our teacher is very good at explaining things clearly.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

53. I worry about giving the wrong answers in this class.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

54. Schoolwork is very hard for me.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

STUDENT QUESTIONNAIRE (Continued) C-2

55. This teacher gives us a chance to show what we are good at.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

56. When I'm in trouble I can count on this teacher to help.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

57. This teacher punishes me for things I didn't do.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

58. Sometimes worrying about my schoolwork makes me feel sick.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

59. Our class gets a lot done each day.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

60. Our teacher makes sure that each one of us gets a chance to talk.
SD-STRONGLY DISAGREE D-DISAGREE A-AGREE SA-STRONGLY AGREE

* *

The remaining items concern your reaction to my presence in the classroom as you can best recall. Put an X on the appropriate letter. Each item applies to visits after the first visit.

61. After my first visit, how distracted were you with my presence in the classroom?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

62. To what extent, when I was there, do you think the overall classroom situation differed from what it usually is like?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

63. To what extent, when I was there, did you refrain (fail) to participate in class discussions because of my presence in the classroom?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

64. To what extent, when I was there, do you think the teacher taught differently from the way he or she usually teaches?
(a) Very much (b) Somewhat (c) Slightly (d) Not at all

Be sure every item has been answered. Mail the questionnaire using the envelope provided. Thank you very much. RMS

C-3

MODIFICATIONS OF MICHIGAN STUDENT QUESTIONNAIRE ITEMS

The *Michigan Student Questionnaire* (MSQ) contains 60 items, all of which were used for the *Student Questionnaire* (SQ). Slight changes[1] in wording were made in the following MSQ items:

Item 6. "This teacher lets some *kids* get by without working very hard" was modified to "this teacher lets some students get by without working very hard."

Item 8. "This teacher *lets us all have turns doing the jobs that are fun*" was modified to "this teacher gives all of us opportunities for sharing in enjoyable tasks."

Item 9. "I think this teacher picks on some *boys and girls* unfairly" was modified to "I think this teacher picks on some students unfairly."

Item 16. "This teacher is always fair with each *boy and girl*" was modified to "this teacher is always fair with each student."

Item 18. "This teacher always asks the OTHER *kids* the EASY questions" was modified to "this teacher always asks the OTHER students the EASY questions."

Item 25. "This teacher really understands *boys and girls* my age" was modified to "this teacher really understands students my age."

Item 36. "The *children* in this class behave well without being told" was modified to "the students in this class behave well without being told."

Item 42. "This teacher doesn't listen to what SOME *boys and girls* have to say" was modified to "this teacher doesn't listen to what SOME students have to say."

Item 48. "This teacher likes *children*" was modified to "this teacher likes teen-agers."

Item 51. "This teacher makes sure no *children* get left out of things" was modified to "this teacher makes sure no students get left out of things."

Items 61-64 are the additions of this investigator.

---

[1] The underlined words of the original statement are those changed.

MEMORANDUM ON STUDENT QUESTIONNAIRE C-4

To: Participating Teachers

From: Ray Snider

Re: Nature and Administration of Student Questionnaire

This questionnaire is a slightly modified version of the Michigan Student Questionnaire; in essence, it is the same thing as the Minnesota Student Attitude Inventory. It was modified by me but originally designed by Ned A. Flanders and his associates in their studies of interaction analysis in the classroom. The modifications are to the extent of the addition of Items 61-64 and slight changes in the wording of a few of the statements. Its purpose is to provide an indication of your physics students' attitudes toward their physics class (the class I have been observing), toward you, and toward school work in general. It will give an indication of how the student perceives the classroom situation. It will NOT indicate what you are ACTUALLY DOING; I am visiting your class seven times to get that indication.

A. You will find the following materials enclosed with this Memorandum: (1) One Questionnaire for each of the students currently enrolled in the physics class which I have been observing; (2) For each questionnaire, a stamped and addressed envelope with which the student will mail the completed questionnaire back to me; and (3) A post card with which you can acknowledge receipt and disbursement of these materials. An additional copy of the questionnaire is included for your files.

B. PLEASE DO THE FOLLOWING:

1. Give one copy of the questionnaire and a return envelope to each of your students currently enrolled in the class I have been observing. Do this in the last three or four minutes of your physics class. Students are to take the questionnaire out of the classroom, complete it in their own privacy and mail it back to me.

2. Do not say anything about the nature of the questionnaire as described in the first paragraph of this Memorandum. Tell the students that the questionnaire is connected with the research that I have been doing and urge them to: (a) read the directions carefully; (b) complete and mail it as soon as possible; and (c) give it their serious attention.

MEMORANDUM ON STUDENT QUESTIONNAIRE (Continued)

3. Point out to your students that I have assured you that their responses: (a) will not be seen by you or any other person connected with your school and so will be completely confidential; and (b) will not in any way jeopardize you, them, or their school.

4. Please check the post card and return it to me. If additional questionnaires are needed, please indicate. You may destroy any extras.

5. Please follow the above explicitly so that I may be assured of uniformity of procedures, from class to class, in administering the questionnaires.

C. I AM, AS ALWAYS, VERY GRATEFUL FOR YOUR CONTINUED ASSISTANCE IN THIS PROJECT. THANK YOU VERY MUCH.

Ray M. Snider
Principal Investigator

APPENDIX D

SUMMED INTERACTION MATRICES OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR THE YEAR FOR EACH OF SEVENTEEN REGENTS PHYSICS TEACHERS

Tables D-1 through D-17

TABLE D-1

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 01[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	10 2	0 0	0 0	1 0	5 1	1 0	0 0	0 0	0 0	4 1
2) T Praises or Encourages	0 0	7 2	0 0	0 0	3 1	0 0	0 0	5 1	0 0	6 1
3) T Accepts or Uses S Idea(s)	0 0	1 0	79 19	12 3	21 5	1 0	0 0	2 0	21 5	5 1
4) T Asks Questions	0 0	1 0	0 0	50 12	13 3	0 0	0 0	50 12	18 4	57 13
5) T Lectures	7 2	1 0	3 1	63 15	2470 580	20 5	2 0	0 0	93 22	87 20
6) T Gives Directions	0 0	0 0	0 0	2 0	14 3	76 18	0 0	0 0	8 2	24 6
7) T Criticizes	0 0	0 0	1 0	0 0	12 3	0 0	16 4	1 0	6 1	3 1
8) S Talk--Response	0 0	1 0	20 5	10 2	24 6	0 0	5 1	25 6	2 0	12 3
9) S Talk--Initiation	3 1	0 0	33 8	7 2	94 22	7 2	15 4	0 0	147 35	57 13
10) Silence or Confusion	1 0	10 2	6 1	44 10	90 21	19 4	1 0	16 4	68 16	256 60
Total	21 5	21 5	142 33	189 44	2746 645	124 29	39 9	99 23	363 85	511 120

[a]Total tallies 4,255--based on 210 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-2

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 02[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	19 3	0 0	0 0	2 0	10 1	2 0	0 0	0 0	2 0	1 0
2) T Praises or Encourages	0 0	22 3	2 0	5 1	10 1	0 0	0 0	21 3	13 2	11 1
3) T Accepts or Uses S Idea(s)	1 0	7 1	133 18	51 7	72 10	8 1	3 0	5 1	24 3	13 2
4) T Asks Questions	2 0	7 1	1 0	232 31	37 5	5 1	4 1	231 31	71 9	105 14
5) T Lectures	11 1	9 1	12 2	172 23	3352 447	65 9	19 3	6 1	159 21	178 24
6) T Gives Directions	2 0	0 0	0 0	14 2	68 9	393 52	3 0	0 0	29 4	45 6
7) T Criticizes	0 0	1 0	1 0	8 1	22 3	7 1	49 7	2 0	11 1	12 2
8) S Talk--Response	0 0	10 1	83 11	64 9	72 10	12 2	10 1	12 2	9 1	23 3
9) S Talk--Initiation	1 0	9 1	75 10	65 9	164 22	21 3	14 2	0 0	281 37	86 11
10) Silence or Confusion	0 0	19 3	10 1	82 11	176 23	41 5	11 1	18 2	117 16	237 32
Total	36 5	84 11	317 42	695 93	3983 531	554 74	113 15	295 39	716 95	711 95

[a] Total tallies 7,504--based on 393 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-3

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 03[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	1 0	3 1	0 0	0 0	1 0	0 0	0 0	0 0	1 0	0 0
2) T Praises or Encourages	2 0	21 5	2 0	5 1	7 2	1 0	0 0	21 5	16 4	9 2
3) T Accepts or Uses S Idea(s)	0 0	10 2	163 38	57 13	97 23	0 0	1 0	4 1	22 5	5 1
4) T Asks Questions	0 0	6 1	0 0	178 42	28 7	1 0	0 0	181 43	55 13	104 24
5) T Lectures	1 0	8 2	9 2	207 49	1803 424	18 4	4 1	9 2	59 14	45 11
6) T Gives Directions	0 0	1 0	0 0	3 1	13 3	84 20	2 0	0 0	5 1	28 7
7) T Criticizes	0 0	1 0	0 0	5 1	10 2	3 1	11 3	2 0	4 1	4 1
8) S Talk--Response	0 0	5 1	102 24	36 8	79 19	5 1	9 2	34 8	1 0	9 2
9) S Talk--Initiation	2 0	11 3	80 19	18 4	65 15	2 0	12 3	0 0	134 31	10 2
10) Silence or Confusion	0 0	18 4	3 1	44 10	60 14	22 5	1 0	29 7	37 9	85 20
Total	6 1	84 20	359 84	553 130	2163 508	136 32	40 9	280 66	334 79	299 70

[a] Total tallies 4,254--based on 221 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-4

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 04[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	5 1	1 0	0 0	2 0	2 0	0 0	1 0	1 0	0 0	1 0
2) T Praises or Encourages	1 0	25 5	7 1	10 2	14 3	0 0	2 0	29 6	33 6	17 3
3) T Accepts or Uses S Idea(s)	0 0	11 2	71 14	55 11	51 10	1 0	2 0	5 1	32 6	10 2
4) T Asks Questions	2 0	9 2	1 0	207 40	32 6	2 0	6 1	201 39	97 19	150 29
5) T Lectures	1 0	12 2	0 0	181 35	1525 297	19 4	26 5	8 2	196 38	99 19
6) T Gives Directions	0 0	0 0	0 0	8 2	12 2	63 12	2 0	0 0	13 3	26 5
7) T Criticizes	0 0	2 0	3 1	12 2	43 8	2 0	101 20	8 2	43 8	16 3
8) S Talk--Response	0 0	14 3	83 16	75 15	50 10	1 0	22 4	21 4	6 1	27 5
9) S Talk--Initiation	0 0	36 7	68 13	71 14	225 44	8 2	53 10	1 0	145 28	98 19
10) Silence or Confusion	4 1	28 5	5 1	86 17	113 22	28 5	15 3	25 5	140 27	177 34
Total	13 3	138 27	238 46	707 137	2067 402	124 24	230 45	299 58	705 137	621 121

[a] Total tallies 5,142--based on 258 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-5

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 05[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	10 2	0 0	0 0	5 1	5 1	0 0	0 0	1 0	3 1	4 1
2) T Praises or Encourages	0 0	65 15	4 1	10 2	19 4	0 0	0 0	10 2	34 8	15 3
3) T Accepts or Uses S Idea(s)	2 0	8 2	126 28	28 6	81 18	0 0	1 0	6 1	43 10	17 4
4) T Asks Questions	1 0	6 1	1 0	141 31	25 6	1 0	2 0	121 27	60 13	140 31
5) T Lectures	3 1	12 3	8 2	180 40	1783 398	12 3	7 2	1 0	126 28	87 19
6) T Gives Directions	0 0	2 0	0 0	1 0	6 1	7 2	0 0	0 0	1 0	8 2
7) T Criticizes	0 0	1 0	0 0	10 2	13 3	0 0	23 5	2 0	14 3	4 1
8) S Talk--Response	1 0	3 1	60 13	27 6	46 10	1 0	7 2	12 3	6 1	5 1
9) S Talk--Initiation	4 1	34 8	107 24	42 9	107 24	0 0	24 5	0 0	233 52	39 9
10) Silence or Confusion	7 2	26 6	6 1	54 12	134 30	4 1	3 1	15 3	70 16	98 22
Total	28 6	157 35	312 70	498 111	2219 495	25 6	67 15	168 37	590 132	417 93

[a] Total tallies 4,481--based on 222 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-6

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 06[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	4	1	0	1	2	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0
2) T Praises or Encourages	0	16	1	5	7	2	0	12	11	11
	0	4	0	1	2	0	0	3	3	3
3) T Accepts or Uses S Idea(s)	0	1	89	70	71	1	0	4	34	10
	0	0	21	16	16	0	0	1	8	2
4) T Asks Questions	0	9	1	285	15	1	0	229	92	167
	0	2	0	66	3	0	0	53	21	39
5) T Lectures	1	4	1	173	1470	9	1	8	97	67
	0	1	0	40	339	2	0	2	22	15
6) T Gives Directions	0	0	0	4	5	13	0	0	4	7
	0	0	0	1	1	3	0	0	1	2
7) T Criticizes	0	2	0	6	6	0	3	4	5	3
	0	0	0	1	1	0	1	1	1	1
8) S Talk--Response	0	9	96	92	55	2	12	26	11	11
	0	2	22	21	13	0	3	6	3	3
9) S Talk--Initiation	1	13	90	71	95	3	11	0	262	32
	0	3	21	16	22	1	3	0	60	7
10) Silence or Confusion	2	10	2	92	105	2	2	31	62	87
	0	2	0	21	24	0	0	7	14	20
Total	8	65	280	799	1831	33	29	314	578	395
	2	15	65	184	423	8	7	72	133	91

[a]Total tallies 4,332--based on 217 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-7

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 07[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
2) T Praises or Encourages	0	4	1	6	1	0	0	7	6	7
	0	1	0	1	0	0	0	1	1	1
3) T Accepts or Uses S Idea(s)	0	2	88	157	69	4	1	3	19	36
	0	0	17	31	13	1	0	1	4	7
4) T Asks Questions	0	7	0	463	27	3	2	364	141	274
	0	1	0	90	5	1	0	71	27	53
5) T Lectures	0	2	0	184	780	24	4	6	79	129
	0	0	0	36	152	5	1	1	15	25
6) T Gives Directions	0	0	0	4	10	82	2	18	10	26
	0	0	0	1	2	16	0	4	2	5
7) T Criticizes	0	0	0	11	9	0	19	7	12	8
	0	0	0	2	2	0	4	1	2	2
8) S Talk--Response	0	2	200	120	66	9	18	107	21	17
	0	0	39	23	13	2	4	21	4	3
9) S Talk--Initiation	0	6	85	96	112	10	15	0	278	48
	0	1	17	19	22	2	3	0	54	9
10) Silence or Confusion	0	9	5	240	134	20	5	48	84	258
	0	2	1	47	26	4	1	9	16	50
Total	0	32	379	1281	1208	152	66	560	650	803
	0	6	74	250	235	30	13	109	127	156

[a] Total tallies 5,131--based on 262 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-8

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 08[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	2	0	0	1	4	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0
2) T Praises or Encourages	0	1	1	3	7	0	0	3	10	1
	0	0	0	0	1	0	0	0	1	0
3) T Accepts or Uses S Idea(s)	0	4	141	39	101	2	2	8	84	28
	0	1	21	6	15	0	0	1	13	4
4) T Asks Questions	0	4	0	295	18	0	2	124	63	28
	0	1	0	44	3	0	0	18	9	4
5) T Lectures	4	3	8	131	3227	39	16	2	196	158
	1	0	1	20	481	6	2	0	29	24
6) T Gives Directions	0	1	0	4	30	91	5	2	6	25
	0	0	0	1	4	14	1	0	1	4
7) T Criticizes	0	0	2	6	31	1	52	4	23	18
	0	0	0	1	5	0	8	1	3	3
8) S Talk--Response	0	3	90	14	17	2	13	47	1	11
	0	0	13	2	3	0	2	7	0	2
9) S Talk--Initiation	0	10	161	13	174	4	35	0	407	100
	0	1	24	2	26	1	5	0	61	15
10) Silence or Confusion	1	0	6	28	175	25	12	8	114	176
	0	0	1	4	26	4	2	1	17	26
Total	7	26	409	534	3784	164	137	198	904	545
	1	4	61	80	564	24	20	30	135	81

[a] Total tallies 6,708--based on 335 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-9

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 09[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	0 0	0 0	0 0	1 0	0 0	0 0	0 0	0 0	0 0	0 0
2) T Praises or Encourages	0 0	5 1	0 0	4 1	1 0	0 0	0 0	2 0	1 0	0 0
3) T Accepts or Uses S Idea(s)	0 0	1 0	68 15	73 16	67 15	3 1	0 0	3 1	14 3	27 6
4) T Asks Questions	0 0	1 0	0 0	283 64	21 5	8 2	1 0	180 40	70 16	126 28
5) T Lectures	0 0	2 0	2 0	164 37	1868 420	25 6	1 0	2 0	45 10	141 32
6) T Gives Directions	0 0	0 0	1 0	7 2	16 4	55 12	0 0	1 0	9 2	27 6
7) T Criticizes	0 0	0 0	0 0	3 1	3 1	0 0	25 6	1 0	1 0	4 1
8) S Talk--Response	0 0	0 0	115 26	35 8	54 12	1 0	2 0	12 3	0 0	22 5
9) S Talk--Initiation	0 0	0 0	62 14	16 4	72 16	4 1	2 0	0 0	78 18	23 5
10) Silence or Confusion	1 0	4 1	8 2	104 23	148 33	20 4	6 1	40 9	39 9	219 49
Total	1 0	13 3	256 58	690 155	2250 506	116 26	37 8	241 54	257 58	589 132

[a] Total tallies 4,450--based on 222 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-10

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 10[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	5 1	0 0	0 0	1 0	1 0	0 0	0 0	0 0	0 0	0 0
2) T Praises or Encourages	0 0	2 0	0 0	1 0	1 0	0 0	0 0	1 0	2 0	1 0
3) T Accepts or Uses S Idea(s)	0 0	0 0	40 9	13 3	23 5	0 0	0 0	0 0	7 2	4 1
4) T Asks Questions	0 0	1 0	0 0	107 25	11 3	1 0	0 0	36 8	45 11	82 19
5) T Lectures	2 0	0 0	0 0	71 17	2820 661	4 1	2 0	3 1	36 8	163 38
6) T Gives Directions	0 0	0 0	0 0	0 0	3 1	4 1	1 0	0 0	2 0	4 1
7) T Criticizes	0 0	0 0	0 0	3 1	5 1	0 0	19 4	0 0	5 1	4 1
8) S Talk--Response	0 0	0 0	18 4	7 2	19 4	0 0	1 0	2 0	0 0	3 1
9) S Talk--Initiation	0 0	2 0	26 6	28 7	40 9	3 1	7 2	0 0	100 23	20 5
10) Silence or Confusion	0 0	3 1	3 1	52 12	178 42	2 0	6 1	8 2	29 7	175 41
Total	7 2	8 2	87 20	283 66	3101 727	14 3	36 8	50 12	226 53	456 107

[a] Total tallies 4,268--based on 214 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-11

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 11[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	0 0	0 0	0 0	0 0	1 0	0 0	0 0	0 0	0 0	0 0
2) T Praises or Encourages	0 0	1 0	0 0	0 0	2 0	0 0	0 0	1 0	2 0	4 1
3) T Accepts or Uses S Idea(s)	0 0	0 0	74 11	26 4	79 12	1 0	2 0	1 0	30 5	13 2
4) T Asks Questions	1 0	1 0	0 0	131 20	7 1	0 0	0 0	100 15	48 7	66 10
5) T Lectures	0 0	2 0	2 0	111 17	4376 661	11 2	3 0	3 0	110 17	229 35
6) T Gives Directions	0 0	0 0	0 0	2 0	8 1	37 6	0 0	0 0	1 0	14 2
7) T Criticizes	0 0	0 0	0 0	0 0	11 2	1 0	23 3	1 0	16 2	3 0
8) S Talk--Response	0 0	0 0	72 11	18 3	27 4	0 0	3 0	3 0	0 0	2 0
9) S Talk--Initiation	0 0	0 0	75 11	18 3	104 16	2 0	22 3	0 0	99 15	14 2
10) Silence or Confusion	0 0	6 1	3 0	48 7	232 35	10 2	2 0	16 2	28 4	257 39
Total	1 0	10 2	226 34	354 54	4847 733	62 9	55 8	125 19	334 50	602 91

[a] Total tallies 6,616--based on 317 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-12

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 12[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	20 4	0 0	1 0	5 1	12 2	0 0	0 0	0 0	3 1	0 0
2) T Praises or Encourages	1 0	31 6	1 0	12 2	7 1	0 0	0 0	6 1	11 2	7 1
3) T Accepts or Uses S Idea(s)	4 1	4 1	177 35	78 15	81 16	1 0	4 1	4 1	60 12	18 4
4) T Asks Questions	1 0	11 2	2 0	248 49	15 3	1 0	1 0	171 34	139 28	96 19
5) T Lectures	8 2	7 1	7 1	162 32	1991 395	15 3	4 1	4 1	134 27	99 20
6) T Gives Directions	1 0	0 0	0 0	5 1	8 2	21 4	0 0	3 1	4 1	23 5
7) T Criticizes	0 0	0 0	2 0	4 1	22 4	1 0	31 6	2 0	18 4	3 1
8) S Talk--Response	0 0	3 1	105 21	43 9	36 7	0 0	6 1	17 3	4 1	12 2
9) S Talk--Initiation	3 1	6 1	124 25	60 12	147 29	12 2	30 6	0 0	163 32	58 11
10) Silence or Confusion	3 1	14 3	12 2	68 13	112 22	14 3	7 1	19 4	67 13	88 17
Total	41 8	76 15	431 85	685 136	2431 482	65 13	83 16	226 45	603 120	404 80

[a] Total tallies 5,045--based on 252 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-13

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 13[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	1 0	0 0	1 0	1 0	2 0	0 0	0 0	0 0	1 0	0 0
2) T Praises or Encourages	0 0	3 1	1 0	3 1	2 0	0 0	0 0	5 1	10 2	5 1
3) T Accepts or Uses S Idea(s)	1 0	2 0	123 28	57 13	41 9	3 1	2 0	6 1	98 22	11 3
4) T Asks Questions	0 0	4 1	2 0	164 37	9 2	0 0	1 0	102 23	140 32	66 15
5) T Lectures	1 0	3 1	3 1	80 18	1194 272	14 3	9 2	10 2	178 41	68 15
6) T Gives Directions	0 0	0 0	0 0	4 1	7 2	16 4	3 1	2 0	9 2	18 4
7) T Criticizes	0 0	0 0	1 0	6 1	14 3	2 0	25 6	2 0	29 7	5 1
8) S Talk--Response	0 0	3 1	57 13	37 8	25 6	2 0	6 1	31 7	8 2	10 2
9) S Talk--Initiation	3 1	9 2	150 34	90 20	181 41	12 3	30 7	3 1	467 106	226 51
10) Silence or Confusion	0 0	5 1	6 1	46 10	85 19	10 2	8 2	18 4	231 53	64 15
Total	6 1	29 7	344 78	488 111	1560 355	59 13	84 19	179 41	1171 267	473 108

[a] Total tallies 4,393--based on 217 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-14

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 14[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	1 0	0 0	0 0	0 0	3 1	0 0	0 0	0 0	0 0	0 0
2) T Praises or Encourages	0 0	8 1	1 0	14 3	5 1	0 0	0 0	2 0	17 3	3 1
3) T Accepts or Uses S Idea(s)	0 0	10 2	87 16	65 12	128 24	5 1	3 1	3 1	91 17	26 5
4) T Asks Questions	0 0	2 0	0 0	149 28	5 1	1 0	1 0	100 19	183 34	39 7
5) T Lectures	3 1	9 2	14 3	118 22	2063 384	29 5	7 1	1 0	246 46	123 23
6) T Gives Directions	0 0	0 0	2 0	4 1	17 3	65 12	3 1	5 1	30 6	47 9
7) T Criticizes	0 0	0 0	0 0	2 0	34 6	4 1	27 5	5 1	30 6	12 2
8) S Talk--Response	0 0	4 1	65 12	14 3	10 2	3 1	12 2	12 2	6 1	6 1
9) S Talk--Initiation	0 0	13 2	241 45	69 13	231 43	34 6	52 10	1 0	89 17	123 23
10) Silence or Confusion	0 0	4 1	8 1	45 8	117 22	32 6	9 2	3 1	161 30	160 30
Total	4 1	50 9	418 78	480 89	2613 486	173 32	114 21	132 25	853 159	539 100

[a] Total tallies 5,376--based on 268 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-15

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 15[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	1 0	0 0	0 0	1 0	2 0	0 0	0 0	0 0	0 0	1 0
2) T Praises or Encourages	0 0	6 1	1 0	1 0	2 0	0 0	0 0	15 3	5 1	6 1
3) T Accepts or Uses S Idea(s)	0 0	2 0	120 27	99 22	69 15	3 1	3 1	5 1	55 12	37 8
4) T Asks Questions	1 0	3 1	0 0	214 48	15 3	0 0	0 0	159 36	101 23	211 47
5) T Lectures	2 0	1 0	2 0	149 33	885 198	26 6	2 0	4 1	75 17	166 37
6) T Gives Directions	0 0	0 0	0 0	5 1	18 4	59 13	1 0	1 0	10 2	37 8
7) T Criticizes	0 0	0 0	0 0	8 2	6 1	6 1	4 1	0 0	8 2	4 1
8) S Talk--Response	0 0	7 2	129 29	36 8	20 4	0 0	5 1	6 1	3 1	17 4
9) S Talk--Initiation	0 0	3 1	132 29	54 12	99 22	8 2	20 4	0 0	262 59	124 28
10) Silence or Confusion	1 0	14 3	9 2	137 31	196 44	29 6	1 0	33 7	183 41	333 74
Total	5 1	36 8	393 88	704 157	1312 293	131 29	36 8	223 50	702 157	936 209

[a] Total tallies 4,478--based on 226 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-16

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 16[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	2 0	0 0	0 0	3 1	2 0	0 0	0 0	0 0	0 0	2 0
2) T Praises or Encourages	0 0	13 3	1 0	5 1	2 0	0 0	0 0	11 2	8 2	12 2
3) T Accepts or Uses S Idea(s)	0 0	2 0	123 25	64 13	86 18	11 2	3 1	2 0	66 14	25 5
4) T Asks Questions	1 0	6 1	1 0	189 39	19 4	1 0	1 0	140 29	147 30	190 39
5) T Lectures	3 1	6 1	7 1	211 43	1412 290	32 7	17 3	11 2	126 26	139 29
6) T Gives Directions	0 0	1 0	1 0	15 3	31 6	68 14	6 1	0 0	18 4	32 7
7) T Criticizes	0 0	3 1	0 0	11 2	37 8	11 2	39 8	8 2	41 8	8 2
8) S Talk--Response	0 0	1 0	91 19	35 7	38 8	4 1	13 3	13 3	2 0	7 1
9) S Talk--Initiation	1 0	3 1	150 31	67 14	149 31	17 3	72 15	0 0	115 24	42 9
10) Silence or Confusion	2 0	17 3	8 2	95 19	188 39	28 6	7 1	19 4	93 19	168 34
Total	9 2	52 11	382 78	695 143	1964 403	172 35	158 32	204 42	616 126	625 128

[a] Total tallies 4,877--based on 247 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies.

TABLE D-17

SUMMED INTERACTION MATRIX OF TOTAL OBSERVATIONS IN ALL MAJOR ACTIVITIES FOR YEAR - TEACHER 17[a]

Category	1	2	3	4	5	6	7	8	9	10
1) T Accepts Feeling	2 0	0 0	1 0	1 0	1 0	0 0	0 0	0 0	3 1	0 0
2) T Praises or Encourages	1 0	11 2	4 1	0 0	9 2	0 0	0 0	3 1	9 2	7 1
3) T Accepts or Uses S Idea(s)	1 0	3 1	141 29	39 8	80 16	0 0	1 0	3 1	74 15	26 5
4) T Asks Questions	0 0	0 0	6 1	236 48	25 5	3 1	1 0	53 11	178 36	103 21
5) T Lectures	1 0	8 2	12 2	152 31	1713 349	21 4	5 1	7 1	166 34	151 31
6) T Gives Directions	0 0	0 0	1 0	4 1	21 4	30 6	0 0	0 0	2 0	26 5
7) T Criticizes	0 0	0 0	0 0	2 0	7 1	0 0	6 1	0 0	8 2	5 1
8) S Talk--Response	2 0	1 0	30 6	15 3	22 4	0 0	2 0	4 1	1 0	4 1
9) S Talk--Initiation	0 0	10 2	160 33	92 19	185 38	12 2	12 2	0 0	219 45	130 26
10) Silence or Confusion	1 0	11 2	13 3	64 13	173 35	18 4	1 0	11 2	160 33	180 37
Total	8 2	44 9	368 75	605 123	2236 456	84 17	28 6	81 17	820 167	632 129

[a] Total tallies 4,906--based on 248 minutes of interaction. Upper number is tallies; lower is tallies per thousand, i.e., ten times the per cent of total tallies

APPENDIX E

FORMS AND MEMORANDA ASSOCIATED WITH CPT AND TOUS TESTING

Form E-1	Memorandum Concerning Distribution of Test Materials
Form E-2	Instructions for Testing
Form E-3	Memorandum Concerning Test Scheduling and Reports of Test Results to Teachers
Form E-4	Cover Letter Accompanying Report of Test Results to Teachers
Form E-5	Information Sheet on CPT and TOUS

MEMORANDUM

Form E-1

To: at
From: Ray Snider
Re: S-2103 Fall Testing of CPT and TOUS - Delivery of Materials.

1. Physics class to be studied: _____ students, meeting time __________. I am planning the fall observations of this class on the following dates; please advise if there is an objection: ____________________.

2. Materials sent for fall testing:

   CPT: _____ booklets; _____ answer sheets.
   TOUS: _____ booklets; _____ answer sheets.

   _____ ELECTROGRAPHIC pencils; _____ return envelope.
   Sent by: first class mail ( ); personal delivery ( ).
   To be received by: date __________; time __________.

3. Your testing schedule:
   CPT: week of Oct. 19; TOUS: week of Oct. _____.

4. Materials yet to be sent for fall testing: ______________________________

5. Return of materials:
   a) All test booklets, used answer sheets and pencils must be returned. Unused answer sheets may be kept. Make no attempt to test students who are absent from the initial testing, i.e., have no "make-ups".
   b) We are a bit short in supply of TOUS and must, in some cases, send them back out to teachers who will be giving it during the week of Oct. 26. Hence it is important that those giving TOUS during the week of Oct. 19 return the materials promptly.
   c) In some cases, materials will be mailed back (stamps will be sent in a few days); in others, materials will be picked up by me. In your case: __________

6. Other: ______________________________

Your assistance and cooperation are greatly appreciated. Thank you very much.

In any cases of difficulty, call collect:
Day: 607-275-2379; Night: 607-273-5941.

MEMORANDUM

Form E-2

To: Participating Teachers

From: Ray Snider

Re: Spring Testing of CPT and TOUS - INSTRUCTIONS FOR TESTING

1. General:

(a) The Cooperative Physics Test, Form Z (CPT) should be administered to the class I am studying during the week of May 10. The Test On Understanding Science, Form W (TOUS) should be administered during the week of May 10 or during the week of May 17. In all cases, please administer CPT first. Testing can be done at any convenient time of the day; there are to be no "make-ups" for absentees.

(b) Test booklets, answer sheets and electrographic pencils are provided; an extra copy of the booklet is provided for the teacher's use in administering the test.

(c) Urge students not to mark on test booklets.

(d) Tell students that the test has a serious purpose and that in each case it will in no way affect their grades in your course.

(e) It is important that you administer the test as you would the Regents (adhere strictly to the time limit; do not interpret any items for the students; maintain test security; provide comfortable conditions; etc.).

2. CPT:

(a) Pass out answer sheets and scrap paper. Slide rules may be used; trigonometric tables will not be necessary. Electrographic pencils (provided) are to be used.

(b) Have students complete personal information on Form Z side only; answers will be indicated on this side as well.

(c) Tell students that this test does not necessarily cover all of the topics they have studied, and that it may test on some topics they have not studied, BUT THEY

MEMORANDUM (Continued) Form E-2

SHOULD NEVERTHELESS: (i) do their best; (ii) consider each item; and (iii) AVOID WILD GUESSING.

(d) Pass out test booklets; students do not open them until told to do so.

(e) When ready to begin the test, READ THE DIRECTIONS ON THE TEST BOOKLET TO THE STUDENTS.

(f) Note the time limits of 20 minutes for Part 1 and 20 minutes for Part 2.

3. TOUS:

(a) Pass out answer sheets. No scrap paper nor slide rules will be necessary. Have students complete personal information on top of answer sheet. (Electrographic pencils, provided, are to be used.)

(b) Pass out test booklets; not to be opened until told to do so.

(c) READ DIRECTIONS ON BACK OF TOUS BOOKLET TO STUDENTS.

(d) Allow 40 minutes for the test.

Note: All test results, students' names, scores, etc. will be strictly confidential. Only you will receive a list of scores by name. All other records will be by means of student code numbers and after scoring, answer sheets will be destroyed. It is important that the foregoing instructions be followed closely in order to insure the best possible uniformity of testing.

Ray M. Snider
Principal Investigator

MEMORANDUM

Form E-3

To: at
From: Ray Snider
Date: April 22, 1965
Re: Spring Testing

I am now in the process of making arrangements for the spring testing (post-testing) of Cooperative Physics Test (CPT) and Test On Understanding Science (TOUS). I am asking that you administer both tests to the physics class I have been observing in this study. A memorandum on "Instructions for Testing" is attached.

Please administer CPT during the week of May 10. TOUS may be administered during the week of May 10 or during the week of May 17. IN ALL CASES, CPT MUST BE ADMINISTERED FIRST. If it is at all possible, I would like to have you carry out the following:

(a) If you administered both tests during the same week in the fall, then administer both tests during the week of May 10.
(b) If you administered TOUS during the week following the week in which you administered CPT last fall, then administer TOUS during the week of May 17.

If it is impossible to adhere to either (a) or (b) (depending upon which is appropriate) then please make whatever administration arrangements are necessary; but in all cases, please give CPT first.

I am sending you, parcel post, CPT test booklets, CPT answer sheets, TOUS answer sheets, and electrographic pencils. These materials should arrive no later than April 27. (Please store them securely until they are used.) Some of the TOUS test booklets will have to be shared by two schools, hence I will not send them until I know which week you will be giving TOUS. The enclosed post card will serve this purpose as well as an indication that you received the materials all right. Please complete the post card after you get the above materials and mail it back.

I will advise you at a later date on the method of return of all the test materials. In some cases, provisions will be made for return by mail--postage and return envelopes will be provided by me. In other cases, I will pick up the materials at your school office.

You may be assured that you will receive a complete analysis of the test results. I will provide you with pre-test scores, post-test scores, adjusted gain scores and percentiles (based on national norms) on all your students that took any of the tests. These results will be given you in June so that you may tell your students how well they did on the tests. In addition, the summary report of the study, which you will receive in September, will give not only the distributions of all the scores (17 schools), but also correlations and regression analyses which will show, for example, to what extent these scores can predict Regents results. In short, I will make every effort to provide you with any information associated with the test results that may be of interest or use.

I am very grateful for your continuing cooperation and thank you very much.

Ray Snider
Principal Investigator

Materials sent by parcel post:

_____ CPT booklets; _____ CPT answer sheets; _____ TOUS answer sheets; _____ electrographic pencils.

IF ANY PROBLEMS ARISE, PLEASE CALL PERSON-TO-PERSON AND COLLECT AT:
607-273-5941 (evenings) 607-275-2379 (days)

MEMORANDUM

Form E-4

To: Participating Teachers

June, 1965

From: Ray Snider

Re: CPT and TOUS Test Scores

Dear (teacher's name):

Attached you will find a list of all students in your physics class who took at least one of the following four: Cooperative Physics Test, Form Z (CPT) October 1964 and May 1965 and Test On Understanding Science (TOUS), October 1964 and May 1965 (in one case, CPT is November 1964). There is also an Information Sheet on CPT and TOUS which provides some background on the norms.

Scores reported (to you only) are CPT Scaled Scores and TOUS Total Scores. For TOUS, percentiles based on the student's grade level (grade 11 or 12 as the case may be) are provided. In the case of CPT, percentiles are not listed as the norm group upon which they are based (5,000 11th graders who took physics in 1950) is possibly inappropriate. However, if you wish to have students see how they compared to this group, a table of percentiles is provided on the lower half of the Information Sheet.

I am asking that you withhold students' CPT and TOUS scores until after they have taken the Regents examination. This, I believe, is particularly important in the case of CPT. We cannot anticipate a proper interpretation of these scores on the part of the student. A student who did poorly with one of the tests may become unnecessarily disillusioned or discouraged at a point when his Regents is just a few days away. Similarly, one who did particularly well may become overconfident. For this reason, I urge you to follow my request and avoid any chance of such difficulties by making these test scores available only after the Regents is over.

As always, your cooperation is greatly appreciated. Thank you very much.

Ray Snider
Principal Investigator

INFORMATION SHEET ON CPT AND TOUS Form E-5

Cooperative Physics Test, Form Z (CPT)

Number right minus $\frac{1}{4}$ of the number wrong = raw score. From the raw score, the scaled score is produced. Items omitted are not counted; this reduces the effect of a "time factor."

The Scaled Score makes it possible to look at a student's score in terms of a group of scores that is normally distributed (for all practical purposes).

For 5,000 students who took physics in 11th grade in eastern U. S. in 1950, we have the following: (based on 90 schools)

The mean Scaled Score is 57.0

The standard deviation (scaled score) is 8.9

Thus, 68% of 5,000 students had Scaled Scores between 48 and 66; 95% were between 39 and 75; 99% were between 30 and 84.

(Note: When the test was developed, there was not evidence to show that 11th graders did any worse in physics than 12th graders.)

Scaled Score	Percentile	Scaled Score	Percentile
78	99	56	46
76	98	54	37
74	97	52	29
72	95	50	22
70	93	48	16
68	89	46	11
66	84	44	7
64	78	42	5
62	71	40	3
60	63	38	2
58	54	36	1

Test On Understanding Science, Form W (TOUS)

The test has three areas:
Area I: Understanding about the scientific enterprise.

Area II: Understanding about scientists.

Area III: Understanding about the methods and aims of science.

For these, there are 18, 18 and 24 items, respectively.

INFORMATION SHEET ON CPT AND TOUS (Continued)

Reported percentile scores for 11th graders are based on Grade 11 Norms; for 12th graders, based on Grade 12 Norms.

Grade 11 Norms:

985 students who were beginning 11th grade, had biology, and were starting chemistry.

Mean Total Score: 31.57
Standard Deviation: 7.02
Year: 1960

Grade 12 Norms:

742 students who were beginning 12th grade, had biology and chemistry, and were starting physics.

Mean Total Score: 32.25
Standard Deviation: 7.38
Year: 1960

The Total Score reliability is 0.76 (Kuder-Richardson).